LINEAR PROGRAMMING AND APPLICATIONS:

A COURSE TEXT

W. McLewin

Department of Mathematics
University of Manchester

LINEAR PROGRAMMING AND APPLICATIONS

This book is No. 4 in our Monograph Series, the previous titles being:

No. 1: PRODUCTIVITY TRENDS IN A SECTORAL MACRO-ECONOMIC MODEL;
W. F. Gossling;

No. 2: ESTIMATING & PROJECTING INPUT-OUTPUT COEFFICIENTS, edited by
R. I. G. Allen & W. F. Gossling;

No. 3: CAPITAL COEFFICIENTS & DYNAMIC INPUT-OUTPUT MODELS, edited by
W. F. Gossling.

The I.-O.P.C. Conference Series includes:

No. 2: INPUT-OUTPUT & THROUGHPUT, edited by
W. F. Gossling;

No. 3: MEDIUM-TERM DYNAMIC FORECASTING, edited by
W. F. Gossling;

No. 4: INPUT, OUTPUT, & MARKETING, edited by
S. J. Gielnik & W. F. Gossling.

The Occasional Papers of the Input-Output Research Association are as follows:

No. 0: "A New Look at Accountancy in the Firm,"
by T. E. Heywood;

No. 1: "Combinatorial Theory of Demand,"
by S. N. Afriat;

No. 2: "Some Productive Consequences of Engel's Law,"
by W. F. Gossling;

No. 3: "A Marketing Review,"
by J. B. Bretscher.

LINEAR PROGRAMMING AND APPLICATIONS

A

Course

Text

by

Will McLewin

INPUT-OUTPUT PUBLISHING COMPANY

3 WYNDHAM PLACE LONDON W1H·1AP ENGLAND

First published 1980

Reprinted 1981

ISBN Manchester Student Edition: 0-904870-13-8
ISBN International Student Edition: 0-904870-12-X
ISBN Case-bound 'hardback' Edition: 0-904870-11-1

CONTENTS

PREFACE

Linear programming has become established as a standard topic in mathematics degree courses, and it is often a part of mathematics courses for students of economics, computer science, business studies and other disciplines. It is one of the basic subjects included under the general title of Operations Research. The usefulness of linear programming in the real world and the intrinsic interest of the situations which lead to linear programming problems make it an attractive area of study for many students, especially as the mathematical prerequisites (basic linear algebra) are modest.

This book is intended as an undergraduate text for a course of about 40 classes (i.e. lectures and tutorials or examples classes), or as a self-paced reading course with very much less teacher-student contact. The main parts of the first seven chapters, which develop the simplex method, duality and versions of the revised simplex method, can be used for a shorter course on general linear programming of about 15–20 classes. The background linear algebra needed consists of a knowledge of matrices, row and column vectors, their elementary properties and techniques of manipulation, the ideas of linear dependence, bases, matrix-inverses, rank, partitioned matrices and solving systems of linear equations. This is all standard material in a first linear algebra course and is available in many textbooks, for example references {1}, {2}, {3}. There is a good case for the view that linear programming should be included in a first linear algebra course, because it provides an interesting context for the practice and application which are necessary to understand fully the ideas and techniques of elementary linear algebra. However a first algebra course is not usually a suitable situation for discussing the practical problems involved, and there are valuable benefits (frequently overlooked) which can be realised when linear programming is treated as a mathematical topic in its own right.

Linear programming is a subject in which the conceptual and manipulative difficulties, although substantial, allow other qualitative ideas to be examined and emphasised. One of these, the increased familiarity and competence with matrix operations, has already been implied, but there are important differences between linear algebra "in theory" and linear algebra "in practice", and linear programming provides a good context in which to introduce these differences. These

considerations also lead to the distinction between a "method" and an "algorithm", where the first may need human intelligence, commonsense and initiative, but the second must not require any of these talents and must be precise and complete and suitable for conversion into a computer program. In general the methods developed in the book are not referred to or described as algorithms, but an awareness of the distinction has significantly influenced the way the methods are developed and presented.

The presentation also stresses the underlying mathematical structure. This emphasis is a particularly important feature, especially for non-mathematics students, to ensure that the methods do not become just a collection of rules; it also means that studying the material involves development of a mathematical approach and of mathematical maturity.

The contents are based on a 30-lecture course I have given several times at the University of Manchester, with a small amount of extra material. The course has been attended mainly by mathematics students but also by students of other disciplines, and their reactions have greatly influenced the choice of material and the various emphases. In particular I have tried to meet the needs of average students and to provide an accessible rather than a formal and strictly scholarly presentation of the material. This approach does not really handicap the dedicated mathematicians and in my experience it often results in a more beneficial experience for most students (from a general mathematical education point of view). There is no resulting loss of rigour.

It has become fashionable to disparage the tableau approach to the simplex method and to favour a "modern" treatment based on matrix operations. Certainly the tableau approach is more rudimentary, but it does not necessarily lead to less insight. A more sophisticated mathematical technique generally requires more sophisticated mathematical experience in order to use it successfully, although its use can itself facilitate such experience. I believe that both approaches have significant advantages and it is worth spending the extra time required to study both. For this reason the tableau operations of chapter 3 are interpreted as matrix operations at the end of chapter 3 but retained in chapters 4 and 5. Then the simplex method is reviewed and the duality theorem re-proved in chapter 6, using the matrix operations approach. Chapter 7 also uses the matrix operations approach.

As I have aimed at providing a course textbook I am interested in knowing the views of students and teachers who use it, and I would be grateful to anyone who can find the time to write and let me know their reactions and to suggest improvements.

It is customary for the author of a book to express his gratitude to the people who have helped him to write it. In my case many friends, colleagues and acquaintances come to mind who have no direct involvement with this book, but for whose very existence I am grateful and without whom life would be much less enjoyable. William Gossling and Christopher Baker encouraged me to write it but bear no responsibility for its shortcomings, nor does Len Freeman, who helped to minimise them.

Will McLewin
Department of Mathematics,
Manchester University,
Manchester M13 9PL.
England

June 1980.

Notation and Abbreviations

Matrices are denoted by upper-case letters in bold type, for example $\mathbf{A}$, and the element in the i-th row and j-th column is denoted by a_{ij}, the corresponding lower-case letter with suffices i and j, or by $(\mathbf{A})_{ij}$.

All vectors are column vectors, and are denoted by lower case letters in bold type, for example $\mathbf{x}$, and the elements of an n-vector $\mathbf{x}$ by $x_1, x_2, \ldots, x_n$, so that $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Row vectors are column vectors transposed, and are denoted by an upper suffix T: for example $\mathbf{y}^T = (y_1, y_2, \ldots, y_m)$ is a row vector in m-space.

Unit matrices are denoted by $\mathbf{I}$, sometimes with a suffix to indicate the size, so that $\mathbf{I}_m$ is the $m \times m$ unit matrix.

The unit vectors which are the columns of a unit matrix are denoted by $\mathbf{e}_1, \mathbf{e}_2, \ldots$, and the vector $(1, 1, \ldots, 1)^T$ by $\mathbf{e}$. The column vector in m-space which is the j-th column of the $m \times n$ matrix $\mathbf{A}$ is denoted by $\mathbf{a}_{*j}$, $j = 1, 2, \ldots, n$, and the row vector in n-space which is the i-th row of $\mathbf{A}$ is denoted by $\mathbf{a}_{i*}$, $i = 1, 2, \ldots, m$.

Partitioned matrices or vectors are denoted thus $(\mathbf{A}_1, \mathbf{A}_2)$ when partitioned column-wise and $\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}$ when partitioned row-wise. This is also the notation when either of the matrices $\mathbf{A}_1$, $\mathbf{A}_2$ consists of a single column or row, for example $(\mathbf{A}, \mathbf{b})$ is the $m \times (n+1)$ matrix whose $(n+1)$-th column is the vector $\mathbf{b}$.

As already implied, the set of all vectors $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in n-dimensional vector space is referred to simply as n-space.

The rank of a matrix $\mathbf{A}$ is denoted by $r(\mathbf{A})$.

Wherever possible the general element of a vector in m-space has the suffix i and the general element of a vector in n-space has the suffix j.

The objective function, typically $\mathbf{c}^T\mathbf{x}$, is denoted by $f(\mathbf{x})$ or simply by f. When convenient we denote an optimum solution of a linear programming problem by $\mathbf{x}_{opt}$, and the corresponding value of f, $f(\mathbf{x}_{opt})$, by f_{opt}.

We also note here that, for example, $\mathbf{x}$ is used to denote both the *name* of a vector and the *value* (i.e., the n actual numbers of which the vector consists in any particular example). This does not cause confusion and is common practice.

The notation $\mathbf{A} \supset \mathbf{I}_m$ has a special meaning which is defined on page 21.

The following abbreviations are used:

l.p.p. for "linear programming problem",
e.c.c. for "equivalent cost coefficient", see page 28
b.f.s. for "basic feasible solution", see page 18
w.l.o.g. for "without loss of generality".

Each chapter is divided into convenient sections: for example 5.3.5 denotes the fifth section of chapter three. The appropriate section number appears at the top of each page. Numbered equations or expressions, for example (6), begin with (1) in each section and are referred to simply by that number in the same section or in full if in another section, for example (6) of section 3.5.3

The references listed on pages 210–11 are indicated in the text thus {4}.

Theorems are numbered consecutively throughout the text and the sections in which they appear are listed on page 209. The symbol ■ is used to denote the end of the statement of a theorem, and the end of the proof.

In several places, statements in the text which may or may not require some explanation have been left as exercises for the reader (*ER*).

There are many different names and notations used to describe the features of linear programming problems, particularly when they are discussed in the context of a specific application. Alternatives are mentioned at appropriate places in the text whenever particular names are defined.

There are two other points of notation, more literary than mathematical, which should be mentioned here. The use of *optimal* and *optimum* as appropriate seems to cause more confusion than the possible ambiguity of meaning that it avoids. The use of *optimum* as an adjective is ~~now~~ nowadays acceptable so *optimum* is used throughout and *optimal* appears only in *optimality*. In chapter 13, on game theory, the distinction between *strategy* and *stratagem* is maintained although it is a common practice not to do so.

CHAPTER 1

A SORT OF INTRODUCTION

1.1

Linear programming is concerned with the problem of finding the optimum (maximum or minimum) value of a linear function subject to a number of linear constraints on the variables. It is a particular case of the general mathematical optimisation problem in which the objective function and the constraint functions may be non-linear. The general problem can properly be regarded as a branch of mathematical analysis, involving the calculus of functions of many variables. The methods for solving the problem are iterative, and use the ideas of convergence and rate of convergence. There are many different methods which are more or less satisfactory, depending on the particular functions involved. In the linear case, one method (the simplex method) can be used to solve any problem in a finite number of steps. However there are different versions or special methods which are more efficient for particular linear programming problems, and there is the method of ellipsoids discussed in chapter 9.

The use of the word *programming* in this context, and of *mathematical programming* for general optimisation problems, should not be confused with computer programming, although in practice non-trivial problems would be solved on a computer. Many linear programming problems are directly related to real-life situations and the solution of each describes the optimum arrangement or programme for the situation.

We begin by considering briefly two such situations, each of which leads to a classical linear programming problem (*l.p.p.*).

1.2 The Diet Problem

Imagine a dietician who wishes to determine the cheapest possible diet satisfying prescribed nutritional requirements and using certain specified foods.

Let m be the number of nutrients; n the number of foods; b_i, $i = 1, 2, \ldots, m$, the amount of the i-th nutrient required; c_j, $j = 1, 2, \ldots, n$, the cost/unit of the j-th food; and a_{ij} the number of

units of the i-th nutrient in each unit of the j-th food. For convenience we may consider a daily diet and all measurements in grams. If x_j, $j = 1, 2, \ldots, n$, is the number of units of the j-th food in the diet, then the total cost is

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n.$$

The total amount of the i-th nutrient is

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n,$$

which must be greater than or equal to b_i, $i = 1, 2, \ldots, m$. In addition, to rule out macabre possibilities, there are the constraints $x_j \geq 0$, $j = 1, 2, \ldots, n$. Thus the dietician's problem becomes to

choose $x_1, x_2, \ldots, x_n$ such that

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

is a minimum subject to

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n \geq b_i \quad \text{and} \quad x_j \geq 0,$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

We may write this problem as follows:

$$\textit{minimise } f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq 0, \tag{1}$$

where $\mathbf{A}$ is the $m \times n$ matrix of nutrient coefficients, $\mathbf{c}$ the n-vector of cost coefficients and $\mathbf{b}$ the m-vector of nutrient requirements.

1.3 The Transportation Problem

Imagine m sources or depots $D_1, D_2, \ldots, D_m$ where there are $d_1, d_2, \ldots, d_m$ units respectively of some commodity, and n locations or destinations $B_1, B_2, \ldots, B_n$ which require $b_1, b_2, \ldots, b_n$ units of the commodity. The cost of transporting one unit of the commodity from D_i to B_j is c_{ij}. The problem for the person deciding what the transportation arrangements should be is to choose x_{ij}, the amount of the commodity to be transported from D_i to B_j, for $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, such that the total cost $\Sigma_{i=1}^m \Sigma_{j=1}^n c_{ij} x_{ij}$ is a minimum subject to the following constraints:

$$\Sigma_{j=1}^n x_{ij} = \text{total amount taken from } D_i \leq d_i,\ i = 1, 2, \ldots, m,$$
$$\Sigma_{i=1}^m x_{ij} = \text{total amount taken to } B_j \geq b_j,\ j = 1, 2, \ldots, n,$$
$$x_{ij} \geq 0,\ i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n.$$

This problem is again to optimise a linear function of the variables, the x_{ij}, subject to a set of linear constraints.

Suppose that $\Sigma_i d_i = \Sigma_j b_j$. Then in order to satisfy all the destination requirements all of the commodity available at the sources must be

taken and no destination can be supplied with more than its requirement; so in this case we may rewrite the constraints as

$$
\begin{aligned}
\Sigma_{j=1}^{n} x_{ij} &= d_i,\ i = 1, 2, \ldots, m,\\
\Sigma_{i=1}^{m} x_{ij} &= b_j,\ j = 1, 2, \ldots, n, \text{ and}\\
x_{ij} &\geq 0,\ i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n.
\end{aligned}
$$

In matrix notation the problem is

$$\textit{minimise } \mathbf{c}^T\mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} = \tilde{\mathbf{b}},\ \mathbf{x} \geq 0, \tag{2}$$

where

$$
\begin{aligned}
\mathbf{c}^T &= (c_{11}, c_{12}, \ldots, c_{1n}, c_{21}, c_{22}, \ldots, c_{2n}, c_{31}, c_{32}, \ldots, c_{m1}, c_{m2}, \ldots, c_{mn}),\\
\mathbf{x} &= (x_{11}, x_{12}, \ldots, x_{1n}, x_{21}, x_{22}, \ldots, x_{2n}, x_{31}, x_{32}, \ldots, x_{m1}, x_{m2}, \ldots, x_{mn})^T,\\
\tilde{\mathbf{b}} &= (d_1, d_2, \ldots, d_m, b_1, b_2, \ldots, b_n)^T
\end{aligned}
$$

and $\mathbf{A}$ is the $(m + n) \times (mn)$ matrix with the following form (ER):

$$
\begin{array}{r|c|c|c|c|c|}
\hline
 & 1\,1\ldots1 & & & & \\
 & & 1\,1\ldots1 & & & \\
m \text{ rows} & & & 11\ldots1 & & \\
 & & & & \cdots & \\
 & & & & & 1\,1\ldots1 \\
\hline
 & 1 & 1 & 1 & & 1 \\
 & \;1 & \;1 & \;1 & & \;1 \\
n \text{ rows} & \ddots & \ddots & \ddots & \cdots & \ddots \\
 & 1 & 1 & 1 & & 1 \\
\hline
\multicolumn{1}{r}{} & \multicolumn{1}{c}{n \text{ columns}} & \multicolumn{1}{c}{n \text{ columns}} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{n \text{ columns}}
\end{array}
$$

1.4

The diet problem and the transportation problem provide us with several useful insights into the general situation.

Each problem is a mathematical model of a real situation and the solutions are useful only to the extent that the model corresponds to the reality. In the diet problem, for example, the cheapest diet may be too unappetizing to be acceptable: a diet of dried milk powder and soya beans was once suggested for the U.S. army. Also, costs are rarely directly proportional to quantities of commodities, either of buying or of transporting. This aspect of formulating the model

is one we choose to ignore and we just accept some simple-minded interpretations of particular situations which lead to *l.p.p.s.*

We would not expect the coefficient matrix **A** in the diet problem to have any special structure, and it would change for different foods and different identified nutrients. But the coefficient matrix **A** of the transportation problem has a strikingly special structure which will be the same for any transportation problem. We can take advantage of this special, constant structure to devise an efficient method of solution (see chapter 10).

The commodity in the transportation problem may be continuous, for example oil, or discrete, for example pianos. In the latter case the vectors **b** and **d** will have integer elements and we will only be interested in integer solutions. For the transportation problem and some other particular problems this presents no difficulties as we shall see. There is no single method which is suitable for solving integer linear programming problems in general.

1.5

It is useful to consider the case of only two variables, because then the problem can easily be described by a diagram in the (x_1, x_2) plane. The constraint

$$a_1 x_1 + a_2 x_2 = b$$

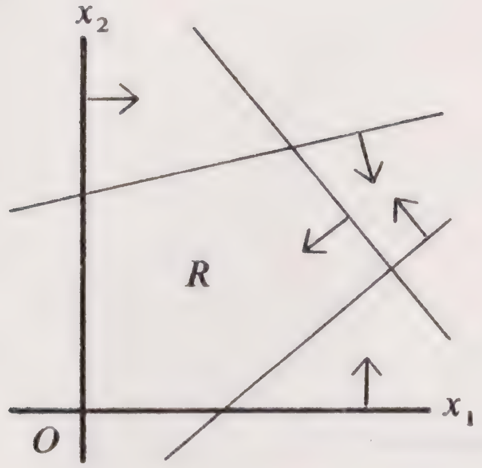
restricts us to a straight line in the (x_1, x_2) plane, and divides the (x_1, x_2) plane into two half-planes: one consists of all points (x_1, x_2) satisfying $a_1 x_1 + a_2 x_2 \geq b$, and the other consists of all points (x_1, x_2) satisfying $a_1 x_1 + a_2 x_2 \leq b$. The same is true, of course, for non-negativity constraints such as $x_1 \geq 0$, $x_2 \geq 0$. The region of the (x_1, x_2) plane satisfying all the constraints for any given problem is thus the intersection of a number of half-planes and there are various possibilities for this region (which we shall call R). These are illustrated in the diagrams on page 5 where a small arrow $\uparrow$ indicates, for each constraint, the half-plane in which the constraint is satisfied.

Now consider the objective function f,

$$f(x_1, x_2) = c_1 x_1 + c_2 x_2.$$

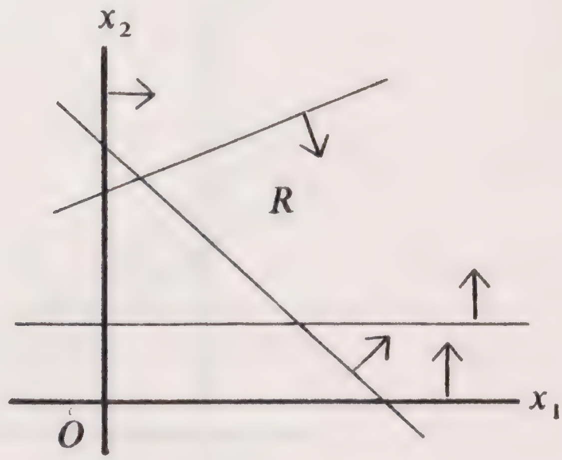
For any two values f_1 and f_2, $c_1 x_1 + c_2 x_2 = f_1$ is a line in the (x_1, x_2) plane, and $c_1 x_1 + c_2 x_2 = f_2$ is another parallel line. In the diagram on page 6, which illustrates the case $c_1 c_2 > 0$, $f_2 > f_1$ if $c_2 > 0$ and $f_2 < f_1$ if $c_2 < 0$ (*ER*). Different values of f correspond to different lines $c_1 x_1 + c_2 x_2 = f$ parallel to the two illustrated.

(i)



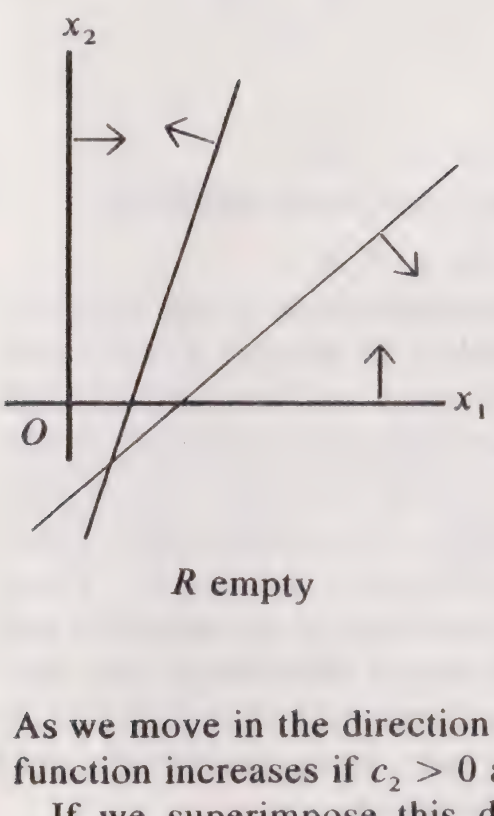
R bounded

(ii)



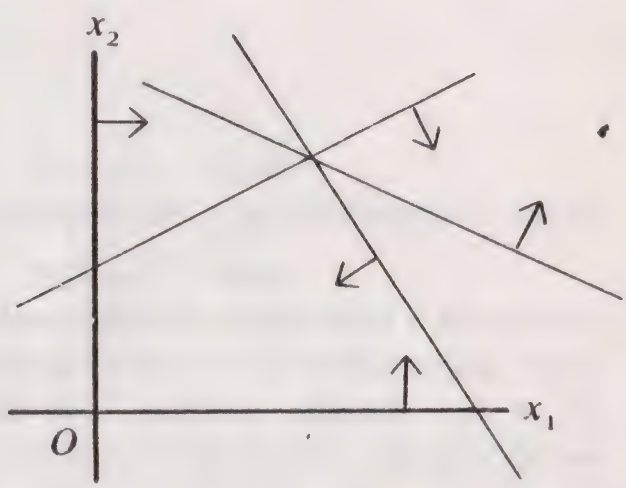
R unbounded

(iii)



R empty

(iv)

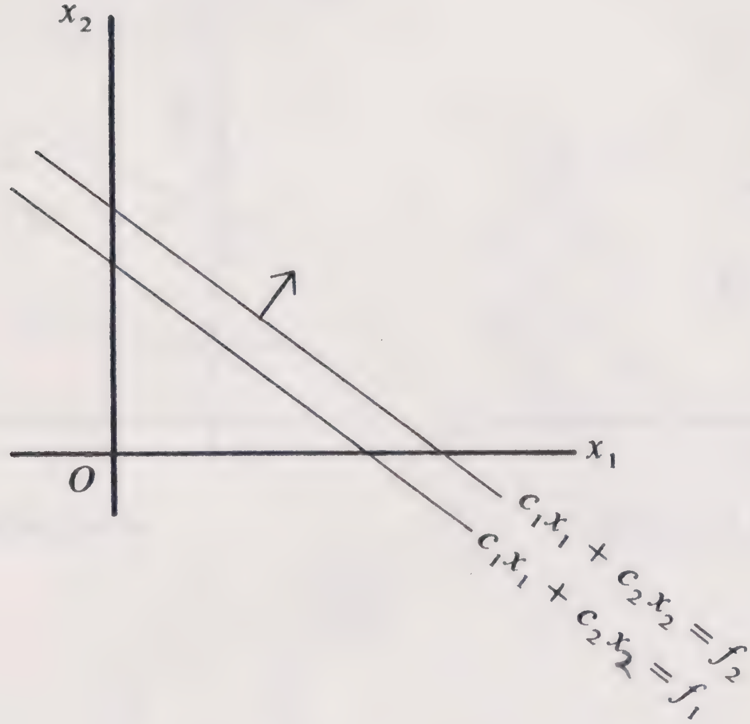


R a single point.

As we move in the direction of the arrow the value f of the objective function increases if $c_2 > 0$ and decreases if $c_2 < 0$.

If we superimpose this diagram on say diagram (i) above, then it is clear that the maximum or minimum value of f is attained on the boundary of R, and either at a vertex of R or along one side of R. Diagrams (ii) and (iii) indicate that there may be no optimum solution for some functions f or no optimum solution for any function f.

For more general problems involving n variables $x_1, x_2, \ldots, x_n$ the situation is similar. The points $\mathbf{x}$ satisfying a constraint



$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$$

lie on a hyperplane in n-dimensional space and those satisfying

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \geq b_1$$

constitute a half-space. An algebraic characterisation of the observations made above for 2-space is established in chapter 2 and leads us to the simplex method. For the moment we just remark that we should not rely on the two-dimensional diagrams too much, useful though they are; in particular, we cannot easily illustrate a non-trivial problem with equality constraints. A similar comment should be made about the numerical examples used to illustrate the theory. These generally have two or three constraints and four to six variables and this makes them just large enough for useful illustration, but they are not really large enough to need the theoretical development. In real-life problems of medium size we may have several hundred variables, and in large problems many thousands.

Exercises 1

1. How is the *l.p.p.* (1) of section 1.2 for the diet problem changed if
   (i) the nutritional requirements are to be satisfied exactly,
   (ii) the nutritional requirements include maximum as well as minimum quantities for some nutrients,
   (iii) instead of minimising the cost, the dietician decides to maximise the attractiveness of the diet, subject to a maximum cost c, by giving each food an enjoyment coefficient p_j, $j = 1, 2, \ldots, n$?
2. For the diet problem, discuss the effect of changing the units. Suppose, for example, that measurement by volume was preferred to measurement by weight for some food.
3. How is the *l.p.p.* (2) for the transportation problem and the optimum solution changed if all transportation costs from a particular depot or to a particular destination are increased by k (>0)?
4. A manager of a company wishes to supply n of the company's factories with specified quantities of a certain raw material. The company advertises its desire to buy this raw material and receives offers of specified amounts and prices from m suppliers. The manager works out the (mn) transportation costs and then has to decide how much to buy from each supplier and which factories to supply with it. Formulate the manager's problem as a *l.p.p.* of transportation type.

   This is a simple version of the contract-award problem. Suggest some likely additional complications in practice.
5. A manufacturer has amounts b_i, $i = 1, 2, \ldots, m$, of m resources which he uses to make n products. He knows the amount a_{ij} of the i-th resource needed to produce one unit of the j-th product, and the profit c_j he makes on one unit of the j-th product. Express as a *l.p.p.* the manufacturer's problem of choosing how much of each product to make, so that his total profit is maximised subject to his available resources. (In this context the elements a_{ij} are *input-output coefficients,* sometimes called *requirement* or *activity coefficients.*)

6. Examine the following *l.p.p.s* graphically. They illustrate the various situations that can occur. In each case $x_1, x_2 \geq 0$.

(i) $3x_1 + 5x_2 \leq 15$
$5x_1 + 2x_2 \leq 10$
maximise $5x_1 + 3x_2$.

(ii) $3x_1 + 5x_2 \leq 15$
$5x_1 + 2x_2 \leq 10$
maximise $2.5\,x_1 + x_2$.

(iii) $x_1 - x_2 \geq -1$
$-x_1 + 2x_2 \leq 4$
maximise $2x_1 + 2x_2$.

(iv) $x_1 + x_2 \leq 1$
$2x_1 + 2x_2 \geq 4$
maximise $3x_1 - 2x_2$.

(v) $x_1 - x_2 \geq 0$
$3x_1 - x_2 \leq -3$
maximise $x_1 + x_2$.

(vi) $-x_1 + x_2 \geq 1$
$x_1 + x_2 \leq 1$
maximise $c_1 x_1 + c_2 x_2$.

7. Discuss the advantages and disadvantages of "simple" mathematical models.
8. By means of a simple diagram in the (x_1, x_2) plane, show that we cannot solve a *l.p.p.* in which an integer solution is required by finding the optimum general solution and then taking the nearest "integer point" to this solution.

NOTES

NOTES

CHAPTER 2

CONVERSION TO SPECIFIED FORM; BASIC, FEASIBLE AND OPTIMUM SOLUTIONS

2.1

Constraints involving $>$ or $<$ do not concern us. Mathematically they define *open* sets of points on which a function may approach arbitrarily close to an optimum value but not actually attain it. In the few cases in which they are appropriate in practice, they can usually be easily replaced by meaningful constraints involving $\geq$ or $\leq$.

As we have seen, the constraints in a *l.p.p.* may involve $\geq$, $\leq$, $=$ or a mixture. We now see, by means of simple examples, how the nature of constraints can be changed by the introduction of extra non-negative variables.

(i) Inequalities can be reversed by multiplying by -1. The inequality constraint

$$-2x_1 + 3x_2 \leq \quad 5$$

is equivalent to the constraint

$$2x_1 - 3x_2 \geq -5.$$

(ii) Inequality constraints can be converted to equality constraints by introducing slack or surplus variables. The constraint

$$2x_1 - 3x_2 \leq 5$$

is equivalent to the two constraints

$$2x_1 - 3x_2 + x_3 = 5, \; x_3 \geq 0.$$

Here x_3 is called a *slack* variable: it tells us how much slack there is before the constraint becomes active or binding.

The constraint

$$2x_1 - 3x_2 \geq 5$$

is equivalent to the two constraints

$$2x_1 - 3x_2 - x_3 = 5, \; x_3 \geq 0.$$

Here x_3 is called a *surplus* variable.

(iii) Equality constraints can be converted into pairs of inequality constraints. The constraint

$$2x_1 - 3x_2 = 5$$

is equivalent to the two constraints

$$2x_1 - 3x_2 \leq 5,\ 2x_1 - 3x_2 \geq 5.$$

Note that an equality constraint cannot be converted into an inequality constraint by the introduction of a slack or a surplus variable (ER).

(iv) A variable not restricted in sign is called a *free variable*, and can be replaced by the difference of two non-negative variables. So the constraints

$$2x_1 - 3x_2 \leq 5,\ \ x_1 + x_2 \leq 1, \qquad (1)$$

where x_1, x_2 are free variables, are equivalent to the constraints

$$\begin{aligned} 2z_1 - 3z_2 + \ \ z_3 &\leq 5, \\ z_1 + \ \ z_2 - 2z_3 &\leq 1,\ \ z_1, z_2, z_3 \geq 0, \end{aligned}$$

where $\qquad x_1 = z_1 - z_3,\ \ x_2 = z_2 - z_3.$

The constraints (1) are also equivalent to the constraints

$$\begin{aligned} 2z_1 - 2z_2 - 3z_3 + 3z_4 + z_5 &= 5, \\ z_1 - \ \ z_2 + \ \ z_3 - \ \ z_4 + z_6 &= 1,\ \ z_1, \ldots, z_6 \geq 0, \end{aligned}$$

where $x_1 = z_1 - z_2$, $x_2 = z_3 - z_4$ and z_5, z_6 are slack variables.

(v) We also note that

$$\underset{x \in R}{maximum}\,(2x_1 - 3x_2) = -\underset{x \in R}{minimum}\,(-2x_1 + 3x_2).$$

2.2

Two particular forms of *l.p.p.* which we shall call *standard form* and *canonical form* are of special interest.

It should be pointed out that it is not unknown for these names to be given different meanings and for these forms to be given different names.

(i) Standard Form

minimise $f(\mathbf{x}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$ *subject to*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &\geq b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &\geq b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &\geq b_m, \ \ \textit{and} \\ x_1, x_2, \ldots, x_n &\geq 0. \end{aligned}$$

In matrix notation, we have

$$\textit{minimise } \mathbf{c}^T\mathbf{x} \textit{ subject to} \qquad (1)$$
$$\mathbf{Ax} \geq \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}.$$

(ii) **Canonical Form**

$$\textit{minimise } \mathbf{c}^T\mathbf{x} \textit{ subject to} \qquad (2)$$
$$\mathbf{Ax} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}.$$

Canonical form is particularly important because *l.p.p.s* are converted to this form before they are solved using the simplex method.

We can always arrange that $\mathbf{b} \geq \mathbf{0}$ for a *l.p.p.* in canonical form without changing the form of the constraints (ER). This is a vital condition for the development of the simplex method, and so whenever we are concerned with solving an *l.p.p.* we include $\mathbf{b} \geq \mathbf{0}$ as part of the definition of canonical form.

2.3

Conversion from standard form to canonical form just requires the introduction of m surplus variables, so that the constraints become

$$\Sigma_{j=1}^{n} a_{ij}x_j - z_i = b_i, \ i = 1, 2, \ldots, m, \qquad (3)$$
$$x_1, \ldots, x_n, \ z_1, \ldots, z_m \geq 0.$$

So *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{Ax} \geq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ becomes

$$\textit{minimise } \tilde{\mathbf{c}}^T\tilde{\mathbf{x}} \textit{ subject to } \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}, \ \tilde{\mathbf{x}} \geq \mathbf{0}, \qquad (4)$$

where $\tilde{\mathbf{c}}^T = (c_1, c_2, \ldots, c_n, 0, 0, \ldots, 0) = (\mathbf{c}^T, \mathbf{0}_m^T)$,
$\tilde{\mathbf{x}}^T = (x_1, x_2, \ldots, x_n, z_1, z_2, \ldots, z_m) = (\mathbf{x}^T, \mathbf{z}^T)$,
$\tilde{\mathbf{b}} = \mathbf{b}$, and

$$\tilde{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & -1 & 0 \ldots & 0 \\ a_{21} & a_{22} & \ldots & a_{2n} & 0 & -1 \ldots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & 0 & 0 \ldots & -1 \end{pmatrix} = (\mathbf{A}, -\mathbf{I}_m).$$

Here $\mathbf{I}_m$ denotes the $m \times m$ unit matrix and $\mathbf{0}_m$ the zero vector with m elements, and we have used the idea of partitioned matrices.

In this case, as an example, the m equations of (3) (and the same m equations of (4)) are the same as

$$(\mathbf{A}, -\mathbf{I}_m)\begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \mathbf{b} = \mathbf{Ax} - \mathbf{I}_m\mathbf{z} = \mathbf{Ax} - \mathbf{z}. \qquad (5)$$

Also, it is worth emphasising that (1) and (4) really are the same

linear programming problem. If $\begin{pmatrix} \mathbf{x}_0 \\ \mathbf{z}_0 \end{pmatrix}$ satisfies the constraints of (4) then $\mathbf{x}_0$ satisfies the constraints of (1), and any $\mathbf{x}_0$ which satisfies the constraints of (1) defines a $\mathbf{z}_0$ such that $\begin{pmatrix} \mathbf{x}_0 \\ \mathbf{z}_0 \end{pmatrix}$ satisfies the constraints of (4). The same argument holds for the vectors at which the minimum is attained.

2.4

From now on we will assume that the *l.p.p.* is in canonical form, with $\mathbf{b} \geq \mathbf{0}$. This can always be achieved by the methods of sections 2.1 and 2.3 although, as exercises 3.3 and 4.6 indicate, it may be more efficient to do something else.

For the moment consider just the equality constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is $m \times n$, and suppose that $m > n$. Either $\mathbf{b}$ is in the column space of $\mathbf{A}$ or it is not. If it is not, then there is no $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$, and $\mathbf{b}$ is not a linear combination of the columns of $\mathbf{A}$, so we do not have an *l.p.p.* to solve because the constraints cannot be satisfied. If it is, then the rank of $\mathbf{A}$, $r(\mathbf{A})$, is the same as the rank of the augmented matrix $(\mathbf{A},\mathbf{b})$, and each is at most n. So at least $(m-n)$ rows of $(\mathbf{A},\mathbf{b})$, i.e. at least $(m-n)$ constraint equations, can be removed because they are linear combinations of the remaining n equations. This takes us to the case $m = n$.

If $m = n$ and $\mathbf{b}$ is not in the column space of $\mathbf{A}$, then $r(\mathbf{A},\mathbf{b}) > r(\mathbf{A})$, and again there is no vector $\mathbf{x}$ satisfying the constraints so we do not have a *l.p.p.* to solve. If $\mathbf{b}$ is in the column space of $\mathbf{A}$, then there is a unique solution $\mathbf{x}$, which is the solution of a corresponding *l.p.p.* in canonical form provided that $\mathbf{x} \geq \mathbf{0}$. If, however, $r(\mathbf{A},\mathbf{b}) = r(\mathbf{A}) = k < n$, then $(n-k)$ equations can be deleted, and this takes us to the case $m < n$.

So we can now assume that $\mathbf{A}$ is $m \times n$ with $m < n$ and we also assume that $r(\mathbf{A}) = m$. These assumptions are a matter of convenience for the development of the simplex method; problems where they are not the case can be dealt with automatically as we shall see in sections 4.3 and 4.5. In practice we do not have to perform the preliminary manipulations that the analysis above implies are necessary to ensure that $m < n$ and $r(\mathbf{A}) = m$.

We have denoted by R the set of vectors satisfying the constraints. Thus

$$R = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}\}.$$

We refer to R as the *feasible region* of n-space and we say a *solution* $\mathbf{x}$ of the equations $\mathbf{Ax} = \mathbf{b}$ is *feasible* if $\mathbf{x} \geq \mathbf{0}$.

A feasible solution for which $\mathbf{c}^T\mathbf{x}$ is a minimum is called an *optimum solution,* and the value $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ of such a solution the *optimum value* (of the *l.p.p.*).

2.5 Convex Sets

A set S is said to be *convex* if $\mathbf{x}_1, \mathbf{x}_2 \in S$, $0<\alpha<1$ implies that $\mathbf{y} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in S$.

The point $\mathbf{y}$ in S is said to be a *convex combination* of $\mathbf{x}_1$ and $\mathbf{x}_2$. (Strictly speaking, $\mathbf{y}$ is a convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ if $\mathbf{y} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ for some α satisfying $0 \leq \alpha \leq 1$, and a *proper,* or nontrivial, convex combination if $0 < \alpha < 1$.)

The set of all convex combinations of $\mathbf{x}_1$ and $\mathbf{x}_2$ is the set of points on the straight-line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$, and so a convex set contains the line segment joining any two points in the set.

A general convex combination of r points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ is any $\mathbf{y}$ where

$$\mathbf{y} = \Sigma_{i=1}^r \alpha_i\mathbf{x}_i, \quad \alpha_i \geq 0, \ i = 1, 2, \ldots, r; \ \Sigma_{i=1}^r \alpha_i = 1,$$

and we can prove, inductively, that if S is convex, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r \in S$ then $\mathbf{y} \in S$. Alternatively, any set of points $\mathbf{x}_1, \ldots, \mathbf{x}_r$ defines a convex set that consists of all points which are convex combinations of them.

A half-space is convex, and so is the intersection of any finite number of convex sets (*ER*). This establishes that R is convex, but we prove this result directly.

Theorem 1

The set R of feasible solutions to a *l.p.p.* is convex■

Suppose $\mathbf{x}_1, \mathbf{x}_2 \in R$. Then $\mathbf{Ax}_1 = \mathbf{b}$, $\mathbf{Ax}_2 = \mathbf{b}$, $\mathbf{x}_1 \geq \mathbf{0}$, $\mathbf{x}_2 \geq \mathbf{0}$.
So $\mathbf{y} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \geq \mathbf{0}$ because for each element y_i of $\mathbf{y}$

$$y_i = \alpha(\mathbf{x}_1)_i + (1-\alpha)(\mathbf{x}_2)_i,$$

which is the sum of non-negative quantities.

Also
$$\begin{aligned}\mathbf{Ay} &= \mathbf{A}(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) = \alpha\mathbf{Ax}_1 + (1-\alpha)\mathbf{Ax}_2 \\ &= \alpha\mathbf{b} + (1-\alpha)\,\mathbf{b} = \mathbf{b}.\end{aligned}$$

Therefore $\mathbf{y} \in R$, and therefore R is convex■

In 2-space, as the diagram (i) on p. 5 indicates, if R is bounded but non-trivial, it is a polygon with no re-entrant vertices and the optimum value of any objective function will be attained at a vertex.

It may possibly be attained at all points of one of the sides of the polygon (see exercise 1.6 (ii)). One can picture the corresponding situation in 3-space where the boundary of R consists of sections of planes. The situation is essentially the same in n-space where the boundary of R consists of sections of $(n-1)$-dimensional hyperplanes and R is called a *polytope.* The picture is more difficult to imagine for a *l.p.p.* in canonical form where the set of solutions of the equality constraints has no interior points. However, in every case the crucial points of R are the vertices, also called extreme points, which are formally defined by their characteristic property as follows.

A point $\mathbf{x}$ of a convex set S is an *extreme point* (or *vertex*) of S if it cannot be written as a proper convex combination of two distinct points of S. At one of the extreme points of R, at least, the optimum value of the objective function is attained. To establish this rigorously using a geometrical approach for *l.p.p.s* in general is rather tedious, partly because of the possibility of R being unbounded, and it would be simpler and sufficient for our purposes to establish the following theorem.

Theorem 2

If a *l.p.p.* has a finite optimum solution, then the optimum value is attained at an extreme point of the feasible region∎

Proving this theorem is the natural next stage in the development of linear programming theory, and the result is a vital step in the development. However, the rather tortuous proof is quite different in spirit from the rest of the development so the proof is not presented here but is to be found in Appendix 1, and the result is established by a different approach in theorem 4.

Another alternative approach is outlined in exercise 2.4. The very reasonable assumptions stated in that exercise can be established with certain provisos, but this is a rather tedious business and the task in exercise 2.4 is just the satisfying end product.

The nature of the proof of theorem 2 indicates that a geometrical approach has severe limitations, and an algebraic approach is needed to actually calculate solutions of *l.p.p.s.* Indeed, the definition of an extreme point indicates this, because although it is clear and succinct there is no obvious way to use it in practice, even to decide whether a particular point of R is an extreme point. So, with the insights provided by the geometrical approach, we now change to an algebraic view of *l.p.p.s* and develop a computable characterisation of extreme

points. This is a crucial step in the development of a computable method or algorithm for solving *l.p.p.s*, and it is worth noting that the essential difference is that from an algebraic point of view the points of R are described with respect to a coordinate system whereas the geometrical approach is coordinate-free.

2.7 Basic Solutions

In canonical form, the equality constraints are $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is $m \times n$ and $m < n$. We are assuming that $r(\mathbf{A}) = m$, so that there are m columns of $\mathbf{A}$ which are linearly independent and without loss of generality (*w.l.o.g.*) we can assume that the first m columns of $\mathbf{A}$ are linearly independent since this could be arranged simply by renumbering the variables $x_1, x_2, \ldots, x_n$. Remember that $\mathbf{Ax} = \mathbf{b}$ is the same as

$$x_1 \mathbf{a}_{*1} + x_2 \mathbf{a}_{*2} + \ldots + x_n \mathbf{a}_{*n} = \mathbf{b},$$

where $\mathbf{a}_{*j}$ is the j-th column of $\mathbf{A}$, for $j = 1, 2, \ldots, n$. If we put $x_{m+1} = x_{m+2} = \ldots = x_n = 0$, then we have a system of m equations in m variables $x_1, x_2, \ldots, x_m$. The matrix of coefficients $\mathbf{A}_1$ is non-singular so there is a unique solution

$$\mathbf{x}_1 = \mathbf{A}_1^{-1}\mathbf{b}, \ (\mathbf{A}_1)_{ij} = a_{ij}, \ i, j = 1, 2, \ldots, m,$$

which gives us a solution $\mathbf{x}$,

$$x_j = \begin{cases} (\mathbf{x}_1)_j, & j = 1, 2, \ldots, m \\ 0, & j = m+1, m+2, \ldots, n, \end{cases}$$

of the constraint equations in which at most m of the variables are non-zero and these non-zero variables correspond to independent columns of $\mathbf{A}$. Another way to write this operation and one we make use of frequently, is to partition $\mathbf{A}$ into $\mathbf{A}_1$, $(m \times m)$, and $\mathbf{A}_2$, $m \times (n - m)$, where

$$\mathbf{A}_1 = \begin{pmatrix} a_{11} & a_{12} \ldots a_{1m} \\ a_{21} & a_{22} \ldots a_{2m} \\ & \vdots \\ a_{m1} & a_{m2} \ldots a_{mm} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} a_{1,m+1} & a_{1,m+2} \cdots a_{1n} \\ a_{2,m+1} & a_{2,m+2} \cdots a_{2n} \\ & \vdots \\ a_{m,m+1} & a_{m,m+2} \cdots a_{mn} \end{pmatrix},$$

and to partition $\mathbf{x}$ into $\mathbf{x}_1$ and $\mathbf{x}_2$ conformally, so that

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \mathbf{x}_1 = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} x_{m+1} \\ \vdots \\ x_n \end{pmatrix}.$$

Then $\mathbf{Ax} = \mathbf{b} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2$; notice that both $\mathbf{A}_1\mathbf{x}_1$ and $\mathbf{A}_2\mathbf{x}_2$ are m-vectors. Putting $\mathbf{x}_2 = \mathbf{0}$ gives $\mathbf{x}_1 = \mathbf{A}_1^{-1}\mathbf{b}$ since $\mathbf{A}_1$ is non-singular.

A solution of the constraint equations in which the non-zero variables correspond to independent columns of $\mathbf{A}$ is called a *basic solution.* At least one basic solution exists if $r(\mathbf{A}) = m$ because there are m independent columns, and in general there are many basic solutions.

The m non-zero variables, $x_1, \ldots, x_m$ above, are called the *basic variables* and the $(n-m)$ zero variables, $x_{m+1}, \ldots, x_n$ above, the *non-basic variables.* The m columns of $\mathbf{A}$ corresponding to the basic variables form a *basis* for m-space and each column of $\mathbf{A}$ can therefore be expressed as a unique linear combination of those m columns.

Of course it may happen that some of the variables in $\mathbf{x}_1$, k say, are zero. In this case we say $\mathbf{x}$ is a *degenerate basic solution.* For a degenerate basic solution, we will regard k of the $(n-m-k)$ variables with value zero as basic variables. Any k of the zero-valued variables can be used provided that the complete set of m basic variables correspond to a non-singular $m \times m$ sub-matrix of $\mathbf{A}$. This suggests, correctly, that a degenerate basic solution is equivalent to a number of coincident basic solutions.

For example, if

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \text{ there are three basic solutions,}$$

$$\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

obtained by putting x_1, x_2, x_3 equal to zero in turn.

If $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ however, we obtain

$$\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

where the last two have x_1 and x_2, and x_1 and x_3 as the basic variables respectively.

2.8 Basic Feasible Solutions

If a basic solution $\mathbf{x}$ of the constraint equations $\mathbf{Ax} = \mathbf{b}$ is non-negative, i.e. $\mathbf{x} \geq \mathbf{0}$, then it is a feasible solution and we have a *basic feasible solution* (*b.f.s.*) of the *l.p.p.* in canonical form. It should

come as no surprise that *b.f.s.s* correspond to extreme points of R, which we now establish in the two halves of theorem 3.

Theorem 3

(i) A *b.f.s.* $\mathbf{x}$ is an extreme point of R■

We assume *w.l.o.g.* that $x_1, x_2, \ldots, x_k > 0$ (and that $x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_m$ are the basic variables although we shall not need this).

Put $\mathbf{x}_1 = (x_1, x_2, \ldots, x_k)^T$, $\mathbf{x}_2 = (x_{k+1}, \ldots, x_n)^T$, and suppose $\mathbf{x}$ is not an extreme point of R. Then there are α, $\mathbf{y}$, $\mathbf{z}$, with $0 < \alpha < 1$, $\mathbf{y} \neq \mathbf{z}$ and $\mathbf{y}, \mathbf{z} \in R$, such that

$$\alpha\,\mathbf{y} + (1-\alpha)\mathbf{z} = \mathbf{x}.$$

This means that $\alpha y_j + (1-\alpha)z_j = x_j$, which for $j > k$ implies $y_j = z_j = 0$. Hence, with $\mathbf{y}_1 = (y_1, y_2, \ldots, y_k)^T$ and $\mathbf{z}_1 = (z_1, z_2, \ldots, z_k)^T$ and $\mathbf{A}_1$ the first k columns of $\mathbf{A}$,

$$\mathbf{A}\mathbf{y} = \mathbf{A}_1\mathbf{y}_1 + \mathbf{A}_2\mathbf{y}_2 = \mathbf{A}_1\mathbf{y}_1 = \mathbf{b} = \mathbf{A}_1\mathbf{z}_1.$$

Hence $\mathbf{A}_1(\mathbf{y}_1 - \mathbf{z}_1) = \mathbf{0}$, which implies that either $\mathbf{y}_1 = \mathbf{z}_1$ or the columns of $\mathbf{A}_1$ are linearly dependent, neither of which is true. Thus $\mathbf{x}$ is an extreme point of R■

(ii) If $\mathbf{x}$ is an extreme point of R, then $\mathbf{x}$ is a *b.f.s*■

We may assume *w.l.o.g.* that $x_1, \ldots, x_k > 0$ and $x_{k+1}, \ldots, x_n = 0$, for some value of k, as in part (i). Partitioning $\mathbf{x}$ and $\mathbf{A}$ as we did in (i), we have

$$\mathbf{A}_1\mathbf{x}_1 = \mathbf{b}, \quad \mathbf{x}_1 > \mathbf{0}.$$

Suppose the columns of $\mathbf{A}_1$ are linearly dependent. Then there is a k-vector $\mathbf{y}_1$ such that

$$\mathbf{A}_1\mathbf{y}_1 = \mathbf{0} \quad \text{and} \quad \mathbf{y}_1 \neq \mathbf{0}.$$

Now for $\lambda > 0$, but sufficiently small,

$$\mathbf{x}_1 + \lambda\mathbf{y}_1 > \mathbf{0},$$
$$\mathbf{A}_1(\mathbf{x}_1 \pm \lambda\mathbf{y}_1) = \mathbf{b},$$

so $\mathbf{z}_1 = \begin{pmatrix} \mathbf{x}_1 + \lambda\mathbf{y}_1 \\ \mathbf{0}_{n-k} \end{pmatrix}$ and $\mathbf{z}_2 = \begin{pmatrix} \mathbf{x}_1 - \lambda\mathbf{y}_1 \\ \mathbf{0}_{n-k} \end{pmatrix}$, both belong to R and $\mathbf{z}_1 \neq \mathbf{z}_2$. However $\mathbf{x} = \frac{1}{2}\mathbf{z}_1 + \frac{1}{2}\mathbf{z}_2$, which implies that $\mathbf{x}$ is not an extreme point of R. This contradicts the definition of $\mathbf{x}$ and so the columns of $\mathbf{A}_1$ are independent. Since $\mathbf{A}$ is $m \times n$ and $r(\mathbf{A}) = m$ we have $k \leq m$, and if $k < m$ then $(m-k)$ of the $(n-k)$ columns of $\mathbf{A}_2$ can be chosen to form a set of m independent columns

(see exercise 2.10). These $(m-k)$ columns of $\mathbf{A}_2$ define the remaining $(m-k)$ basic variables and hence $\mathbf{x}$ is a *b.f.s.* ■

2.9 Theorem 4. The Fundamental Theorem of Linear Programming

If the *l.p.p. minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, where $\mathbf{A}$ is $m \times n$ and $r(\mathbf{A}) = m$, has a feasible solution then it has a *b.f.s.*, and if it has an optimum solution then it has an optimum *b.f.s.* ■

To establish the first assertion, let $\mathbf{x}$ be a feasible solution and assume *w.l.o.g.* that $x_1, x_2, \ldots, x_k > 0$ and $x_{k+1}, x_{k+2}, \ldots, x_n = 0$. If the first k columns of $\mathbf{A}$ are linearly independent then $k \leq m$. If $k = m$ then $\mathbf{x}$ is a *b.f.s.* If $k < m$ then $\mathbf{x}$ is a degenerate *b.f.s.* and $(m-k)$ of the zero elements of $\mathbf{x}$ can be chosen to make up the m basic variables.

If the first k columns of $\mathbf{A}$ are not independent, then there is a k-vector $\boldsymbol{\alpha} \neq \mathbf{0}$ such that

$$\Sigma_{i=1}^{k} \alpha_i \mathbf{a}_{*i} = \mathbf{A}_1 \boldsymbol{\alpha} = \mathbf{0}, \text{ and therefore}$$
$$\Sigma_{i=1}^{k} (x_i - \theta\alpha_i)\mathbf{a}_{*i} = \mathbf{b} \quad \text{for any} \quad \theta.$$

We can arrange that at least one element of $\boldsymbol{\alpha}$ is positive, and so as θ is increased from 0 for some value of θ, θ_0 say, and some s, $1 \leq s \leq k$, $x_s - \theta_0\alpha_s = 0$, $x_i - \theta_0\alpha_i > 0$ for $i \neq s$.

Denoting $x_i - \theta_0\alpha_i$ by x_i', we now have a new feasible solution $\mathbf{x}'$, with at most $(k-1)$ non-zero elements. The process can be repeated until the columns of $\mathbf{A}$ corresponding to non-zero elements of the feasible solution are independent.

The second assertion of theorem 4 is left as an exercise (see exercise 2.11) ■

Noticé that despite the constructive nature of the proof, theorem 4 does not directly provide a method of solving *l.p.p.s* because it does not provide a means of obtaining the feasible or optimum vector $\mathbf{x}$ with which the proof begins. Referring back to section 2.7 we see that indirectly theorem 4 does provide us with a method for solving any *l.p.p.* because it establishes that it is sufficient to find the value of the objective function at all the *b.f.s.s* and choose the optimum of these. Unfortunately this requires us to solve up to $n!/(m!(n-m)!)$ $m \times m$ systems of equations, which is extremely inefficient. In the simplex method in practice, we can expect to do on average *about* as much work as there is in solving two $m \times m$ systems of equations. Even for n, m quite small, $n = 20$, $m = 10$ say, $n!/(m!(n-m)!)$ is rather large, 184,756 in fact.

However, it is now clear that solving a *l.p.p.* only involves solving a non-singular system of equations and the difficulty is that of deciding which system to solve.

We also observe that if the columns of $\mathbf{I}_m$ are present among the columns of $\mathbf{A}$ then we have one *b.f.s.* immediately: namely $x_j = b_i$ if the j-th column of $\mathbf{A}$ is the i-th column of $\mathbf{I}_m$ and $x_j = 0$ otherwise. In particular if we partition $\mathbf{A}$ into $(\mathbf{A}_1, \mathbf{A}_2)$ and $\mathbf{A}_1 = \mathbf{I}_m$, then

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \quad \text{is a } b.f.s.$$

We shall use the notation "$\mathbf{A} \supset \mathbf{I}_m$" ($\mathbf{A}$ contains $\mathbf{I}_m$) to denote that the columns of $\mathbf{I}_m$ are present among the columns of $\mathbf{A}$. This is just a convenient use of notation and does not mean, for us, that each of the columns of $\mathbf{I}_m$ can be written as a linear combination of the columns of $\mathbf{A}$, which is true for any $m \times n$ matrix with rank m.

Exercises 2

1. Convert the following *l.p.p.s* to canonical form:

(i) *maximise* $3x_1 - x_2$ *subject to*
$$2x_1 + 3x_2 \le 6$$
$$x_1 + 7x_2 \ge 4$$
$$x_1 + x_2 = 3$$
$$x_1, x_2 \ge 0.$$

(ii) *maximise* $2x_1 + 3x_3$ *subject to*
$$4x_1 + 2x_2 \le 4$$
$$x_1 + 3x_3 \le 5$$
$$x_1, x_2 \ge 0.$$

(iii) *minimise* $x_1 - x_2$ *subject to*
$$x_1 - x_2 \le 3$$
$$x_1 + x_2 - x_3 = 4$$
$$x_1 \ge 0.$$

2. Repeat section 2.3 but instead convert from canonical form to standard form.
3. Suppose that the optimum solution of a *l.p.p.* in canonical form is not unique. Show that the set of all optimum solutions is convex.
4. For a *l.p.p.* in canonical form, given that R is bounded and has a finite number of extreme points and that any point of R can be written as a convex combination of the extreme points of R, prove that the optimum value is attained at an extreme point.
5. In the diet problem of section 1.2, give a physical interpretation of the fact that the optimum solution will be basic. Give an interpretation of the following possibilities:
   (i) one row of $\mathbf{A}$ is zero,
   (ii) one column of $\mathbf{A}$ is zero,
   (iii) $r(\mathbf{A}) = (m - 1)$.
   Does (iii) imply that the optimum solution will be degenerate? Convert the diet problem to canonical form by introducing m surplus variables. What does it mean if one of them is a basic variable in the optimum solution?
6. Convert the manufacturer's problem, exercise 1.5, to canonical form and give an interpretation of the possibility that a slack variable is basic in the optimum solution. Note that each column of the matrix of coefficients $\mathbf{A}$ describes one of the manufacturer's activities. The slack variables and their corresponding columns in canonical form are called *disposal* activities.
7. Solve the following *l.p.p.s* by finding all *b.f.s.s:*

(i) *maximise* $2x_1 + 3x_2$ *subject to*
$$4x_1 + 2x_2 + x_3 = 4$$
$$x_1 + 3x_2 = 5$$
$$x_1, x_2, x_3 \ge 0.$$

(ii) *minimise* $(1, -1, 1, 1)\mathbf{x}$ *subject to*
$$\begin{pmatrix} 2 & 6 & 2 & 1 \\ 6 & 4 & 4 & 6 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$
$$\mathbf{x} \ge \mathbf{0}.$$

8. Suppose that $\mathbf{x}$ and $\mathbf{y}$ are two distinct non-degenerate extreme point optimum solutions (*b.f.s.s*) of a *l.p.p.* in canonical form. Show that any non-trivial convex combination of $\mathbf{x}$ and $\mathbf{y}$ is an optimum solution but not a *b.f.s.*
9. The unit matrix $\mathbf{I}_3$ appears in the matrix $\mathbf{A}$ for the system of equations below and hence there is an obvious basic solution. By performing suitable row operations on the augmented matrix $(\mathbf{A}, \mathbf{b})$ convert (i) the fourth (ii) the fifth column of $\mathbf{A}$ to a column of $\mathbf{I}_3$ and hence obtain basic solutions in which $\mathbf{x}_4$, $\mathbf{x}_5$ respectively are basic variables.

$$\begin{pmatrix} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 4 & -2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 4 \\ 3 \\ 6 \end{pmatrix}$$

10. Suppose that the $m \times n$ matrix $\mathbf{A}$, with $m < n$, has rank m. Prove that for any set of k independent columns a further $(m-k)$ columns can be found to complete a set of m independent columns. (Just consider the remaining columns one at a time and add them to the set we already have if they are independent.)
11. Prove the second part of the fundamental theorem of linear programming (see section 2.9) by using the same argument that was used to prove the first part and noting that

$$\mathbf{c}^T\mathbf{x} = \Sigma_{i=1}^{k} c_i(x_i - \theta\alpha_i)$$

because $(x_i - \theta\alpha_i)\mathbf{a}_{*i}$ is a feasible solution for θ small enough and positive *or* negative.

NOTES

CHAPTER 3

THE SIMPLEX METHOD

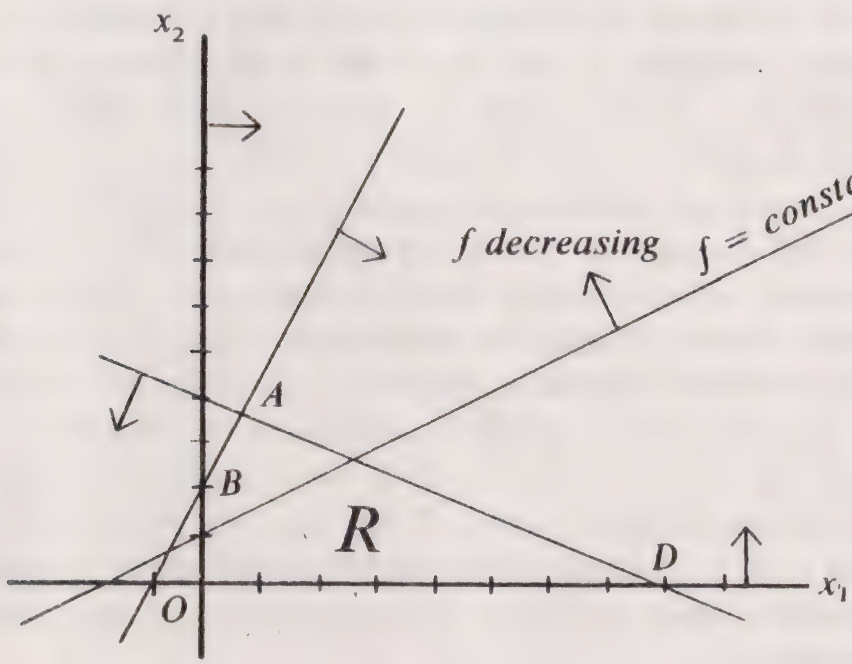
3.1

We first examine a simple example and solve it by a methodical but essentially intuitive process.

Minimise $2x_1 - 3x_2$ *subject to*

$$-2x_1 + \ \ x_2 \le 2,$$
$$x_1 + 2x_2 \le 8, \quad x_1, x_2 \ge 0.$$

This problem is easily solved graphically.



The minimum occurs at A, which is the point $(\frac{4}{5}, \frac{18}{5})$, and there

$$f(\mathbf{x}) = 2x_1 - 3x_2 = -46/5. \tag{1}$$

Converting to canonical form with two slack variables x_3 and x_4 gives

$$\textit{minimise } 2x_1 - 3x_2 + 0x_3 + 0x_4$$

subject to

$$-2x_1 + \ \ x_2 + x_3 = 2, \tag{2}$$
$$x_1 + 2x_2 + x_4 = 8, \ \ x_1, x_2, x_3, x_4 \ge 0, \tag{3}$$

or *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$,

where
$$\mathbf{c}^T = (2,-3,0,0), \tag{4}$$

$$\mathbf{b} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} -2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}. \tag{5}$$

In canonical form we see that $\mathbf{A} \supset \mathbf{I}_2$, so there is one obvious *b.f.s.*: $x_1 = x_2 = 0$, $x_3 = 2$, $x_4 = 8$. This corresponds to the point O in the diagram, and the value of f there, $f_O = 0$.

The three other basic solutions that are feasible are:

$x_1 = x_3 = 0,\ x_2 = 2,\ x_4 = 4,$ the vertex $B, f_B = -6,$

$x_2 = x_4 = 0,\ x_1 = 8,\ x_3 = 18,$ the vertex $D, f_D = 16,$

$x_3 = x_4 = 0,\ x_1 = 4/5\, x_2 = 18/5,$ the vertex $A, f_A = -46/5,$

which confirms that the optimum value of $f(\mathbf{x})$ is attained at A.

Notice that in (1) or (4) we effectively have $f(\mathbf{x})$ expressed in terms of the non-basic variables x_1 and x_2 of the basic solution provided by the columns of $\mathbf{I}_2$ in the matrix $\mathbf{A}$. For any other basic feasible solution one of x_1 and x_2 must be positive, since $x_1 = x_2 = 0$ implies $x_3 = 2$, $x_4 = 8$, so we ask whether increasing x_1 or x_2 from its current value of zero will decrease f. Since the coefficient of x_2 is negative we will decrease f if we increase x_2. By how much can x_2 be increased? The equalities in (2) and (3) must be maintained, so as x_2 is increased, with the other non-basic variable x_1 remaining zero, the basic variables x_3 in the first equation and x_4 in the second must be changed.

For example, in $-2x_1 + x_2 + x_3 = 2$, if we increase x_2 from zero to θ say, we must decrease x_3 by θ, so that the maximum value x_2 can be given is 2, because anything higher would make x_3 negative.

This is possibly easier to follow if we rewrite the equations (2) and (3) as follows,

$$x_3 = 2 + 2x_1 - x_2, \tag{6}$$

$$x_4 = 8 - x_1 - 2x_2, \tag{7}$$

and ask how large can x_2 be made, with $x_1 = 0$, before x_3 or x_4 becomes negative. Equation (6) is the crucial one and gives $x_2 = 2$. However, with $x_2 = 2$ and $x_3 = 0$ we have another *b.f.s.*, with $x_4 = 4$ from (7), and this is the vertex B. The value of $f(\mathbf{x})$ at B is -6 and

$$f_B = -6 = 0 + 2 \times (-3),$$

which is the previous value f_O plus the increase in x_2 multiplied by the coefficient of x_2 in (1).

In order to use the same argument for the new *b.f.s.* we need to convert (1), (2), and (3) to the same form, namely $f(\mathbf{x})$ in terms of non-basic variables x_1, x_3, and equality constraints which express the basic variables in terms of the non-basic variables x_2, x_4. Rearranging (6) and substituting for x_2 in (7) gives

$$x_2 = 2 + 2x_1 - x_3, \quad \text{and} \tag{8}$$

$$x_4 = 8 - x_1 - 2(2 + 2x_1 - x_3),$$

$$\text{i.e.} \quad x_4 = 4 - 5x_1 + 2x_3. \tag{9}$$

Substituting in (1) for x_2 gives

$$f = 2x_1 - 3(2 + 2x_1 - x_3)$$

$$= -6 - 4x_1 + 3x_3. \tag{10}$$

Equations (8) and (9) show clearly what has been done, because the new matrix of coefficients of the equality constraints still contains $\mathbf{I}_2$, but now the columns of $\mathbf{I}_2$ correspond to x_2 and x_4. Also, the cost coefficients in (10) corresponding to basic variables are zero. Thus, putting $x_1 = x_3 = 0$ gives

$$x_2 = 2, \; x_4 = 4, \; f = f_B = -6.$$

From (10) we see that we can further decrease f by increasing x_1 (with x_3 remaining zero), and from (8) and (9) the maximum value we can give x_1 is $4/5$, which makes x_4 zero (and gives $x_2 = 18/5$).

The same rearrangement and substitution gives

$$x_1 = 4/5 + 2/5x_3 - 1/5x_4, \tag{11}$$

$$x_2 = 18/5 - 1/5x_3 - 2/5x_4, \quad \text{and} \tag{12}$$

$$f = -46/5 + 7/5x_3 + 4/5x_4, \tag{13}$$

which corresponds to the vertex A.

Both (all) coefficients of non-basic variables in (13) are positive and the current *b.f.s.* is obtained by giving them the value zero. In any other *b.f.s.* (or any other feasible solution for that matter) at least one of x_3 and x_4 will be positive and so f will be greater; thus the current *b.f.s.* is optimum.

The procedure developed intuitively above is in fact the simplex method. We now go on to describe it formally in general, and to establish an alternative way of describing the operations that are performed. This will, of necessity, appear more complicated, and

it is possible to make the simplex method appear very complicated indeed, so it is important to bear in mind the essential simplicity of the argument and the operations used.

Two comments may help:

(i) The pairs of equations or equality constraints numbered (2) and (3), (6) and (7), (8) and (9), and (11) and (12) are all equivalent. They define the same region of 4-space and hence, together with $\mathbf{x} \geq \mathbf{0}$ and $f(\mathbf{x})$, they define the same *l.p.p.*

The equations $\mathbf{EAx} = \mathbf{Eb}$, i.e. $\mathbf{A}'\mathbf{x} = \mathbf{b}'$, where $\mathbf{A}' = \mathbf{EA}$ and $\mathbf{b}' = \mathbf{Eb}$, are equivalent to the equations $\mathbf{Ax} = \mathbf{b}$ for any non-singular matrix $\mathbf{E}$, and there is a 2×2 matrix, $\mathbf{E}_1$ say, which transforms (2) and (3) to (8) and (9). We did not identify it at the time, but it is in fact

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

(ii) The different expressions for $f(\mathbf{x})$ given by (1), (10) and (13) are also equivalent, and although each has been associated with a particular vertex of R each is valid throughout R and each has the same value for any point of R. The fact that (10) and (13) involve a constant as well as a linear combination of the variables does not make them different from (1). In general, referring back to sections 2.2 and 2.4, we could have assumed that $f(\mathbf{x})$ had the form $\mathbf{c}^T\mathbf{x} + \alpha$, for some constant α, because this constant essentially leaves the *l.p.p.* unchanged.

3.2

We assume that the *l.p.p.* has been converted to canonical form. It is then defined by $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{c}$ and we shall refer to the elements of this $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{c}$ as the *original* coefficients.

The simplex method requires the matrix $\mathbf{A}$ and the vector $\mathbf{c}$ to have a certain form, namely $\mathbf{A} \supset \mathbf{I}_m$ and $c_j = 0$ if x_j is a basic variable. If this is the case, the method may begin at once. If not, some preliminary manipulation is required (which we describe later in sections 4.2–4.5) to produce an equivalent set of constraint equations which we denote using $\mathbf{A}'$ and $\mathbf{b}'$ say, and a set $\mathbf{c}'$ of *equivalent cost coefficients (e.c.c.s.).*

The equivalent cost coefficients $c_1', c_2', \ldots, c_n'$ are often called *relative cost coefficients* or *reduced cost coefficients,* and sometimes *modified cost coefficients* or *simplex criteria.* The name *equivalent*

is preferable since it describes exactly what they are, just as $\mathbf{A}'\mathbf{x} = \mathbf{b}'$ is an equivalent set of equality constraints. Although any particular set $\mathbf{c}'$ of *e.c.c.s* appears associated with or relative to a particular *b.f.s.* the corresponding expression $\mathbf{c}'^T\mathbf{x} - \alpha'$ for $f(\mathbf{x})$ is valid at *all* points of R (see section 3.6). Further comments on nomenclature appear in section 5.7.

One stage of the simplex method produces a new *b.f.s.* and a new set of coefficients, so we shall describe such a stage in terms of

$$\mathbf{A}', \mathbf{b}', \mathbf{c}', (a'_{ij}, b'_i, c'_j, \; i = 1,2, \ldots, m, \quad j = 1,2, \ldots, n) \text{ and}$$
$$\mathbf{A}^*, \mathbf{b}^*, \mathbf{c}^*, (a^*_{ij}, b^*_i, c^*_j, \; i = 1,2, \ldots, m, \quad j = 1,2, \ldots, n).$$

For convenience and *w.l.o.g.* we assume that the first m columns of $\mathbf{A}'$ are the columns of $\mathbf{I}_m$ in order. The situation can then be completely described by the following tableau of coefficients:

$$\begin{array}{|ccccc|cccccc|c|}
\hline
1 & 0 & 0 & \ldots & 0 & a'_{1\,m+1} & a'_{1\,m+2} & \ldots & a'_{1t} & \ldots & a'_{1n} & b'_1 \\
0 & 1 & 0 & \ldots & 0 & a'_{2\,m+1} & a'_{2\,m+2} & \ldots & a'_{2t} & \ldots & a'_{2n} & b'_2 \\
\vdots & & & & & \vdots & & & \vdots & & \vdots & \vdots \\
0 & 0 & & 1 & 0 & a'_{s\,m+1} & a'_{s\,m+2} & \ldots & a'_{st} & \ldots & a'_{sn} & b'_s \\
\vdots & & & & & \vdots & & & \vdots & & \vdots & \vdots \\
0 & 0 & & \ldots & 1 & a'_{m\,m+1} & a'_{m\,m+2} & \ldots & a'_{mt} & \ldots & a'_{mn} & b'_m \\
\hline
0 & 0 & . & \ldots & 0 & c'_{m+1} & c'_{m+2} & \ldots & c'_t & \ldots & c'_n & f + \alpha' \\
\hline
\end{array}\,.$$

We call such a tableau a *simplex tableau.*

Here we have emphasised the t-th column and the s-th row for reasons which will become clear shortly. The current *b.f.s.* is $\mathbf{x}$, where

$$x_j = \begin{cases} b'_j, & j = 1,2, \ldots, m \\ 0, & j = m+1, m+2, \ldots, n. \end{cases}$$

The current basic variables are $x_1, x_2, \ldots, x_m$. The i-th row of the tableau, $i = 1,2, \ldots, m,$ is a shorthand for

$$0x_1 + 0x_2 + \ldots + 0x_{i-1} + x_i + 0x_{i+1} + \ldots + a'_{i\,m+1}x_{m+1} \\ + a'_{i\,m+2}x_{m+2} + \ldots + a'_{it}x_t + \ldots + a'_{in}x_n = b'_i, \quad (1)$$

and the last row of the tableau which we will refer to as the '*c-row*' of the tableau is a shorthand for

$$0x_1 + 0x_2 + \ldots + 0x_m + c'_{m+1}x_{m+1} + c'_{m+2}x_{m+2} + \ldots \\ + c'_t x_t + \ldots + c'_n x_n = f(\mathbf{x}) + \alpha'. \quad (2)$$

Since $x_{m+1} = x_{m+2} = \ldots = x_n = 0$ for the (current) corresponding *b.f.s.* the value of this *b.f.s.* is $-\alpha'$ and in practice, when the symbolic

coefficients are replaced by numbers, the f can be omitted and just the value of α' placed in the bottom right-hand corner.

One stage of the simplex method thus proceeds as follows:

If $c'_j \geq 0$, $j = m+1, m+2, \ldots, n$, the current *b.f.s.* is optimum, since at present $x_{m+1}, x_{m+2}, \ldots, x_n$ are zero and in any different *b.f.s.* at least one will be positive and thus f can only be increased. If at least one *e.c.c.* is negative choose t such that $c'_t = \min\limits_{j=m+1,m+2,\ldots,n} c'_j$. Now f can be decreased by increasing x_t from its current value of zero. Either (i) $a'_{it} \leq 0$, $i = 1, 2, \ldots, m$,

or (ii) $a'_{it} > 0$ for some i, $1 \leq i \leq m$.

(i) If x_t is increased to θ say ($\theta > 0$), then in the i-th constraint, equality is preserved by adding $-a'_{it}\theta$ to x_i for $i = 1, 2, \ldots, m$, and

$x_i(\theta) = x_i - a'_{it}\theta \geq 0$, $i = 1, 2, \ldots, m$, so $\mathbf{x}(\theta)$ defined by

$$(\mathbf{x}(\theta))_j = \begin{cases} x_j(\theta) & j = 1, 2, \ldots, m \\ 0 & j > m, j \neq t \\ \theta & j = t \end{cases} \tag{3}$$

is a feasible solution for any $\theta > 0$. However, the value of this solution $f(\mathbf{x}(\theta))$ is

$$-\alpha' + c'_t\theta, \tag{4}$$

which can be made less than *any* chosen number K say by choosing θ sufficiently large.

Therefore $c'_t < 0$, $a'_{it} \leq 0$, $i = 1, 2, \ldots, m$, implies that the *l.p.p.* has no minimum solution—the values of feasible solutions are unbounded below!

(ii) If x_t is increased to θ say ($\theta > 0$), then in the i-th constraint, equality is preserved by adding $-a'_{it}\theta$ to x_i for $i = 1, 2, \ldots, m$, but now $-a_{it}\theta < 0$ for some i so the maximum value that we can choose for θ is

$$\min_{\substack{i=1,2,\ldots,m \\ a'_{it}>0}} \frac{x_i}{a'_{it}} = \frac{x_s}{a'_{st}} \quad \text{say,}$$

which is equal to

$$\frac{b'_s}{a'_{st}} = \min_{\substack{i=1,2,\ldots,m \\ a'_{it}>0}} \frac{b'_i}{a'_{it}}. \tag{5}$$

If the minimum is attained for several values of i, to make the procedure definite we choose, at present, the minimum of these values for s.

Thus x_s becomes non-basic with value zero, x_t becomes basic with value θ, and other basic variables remain basic but their values change. The value of f decreases by the positive amount $-\theta \times c'_t$, or increases by minus this amount, which is the positive quantity

$$\frac{-b'_s}{a'_{st}} \times c'_t. \tag{6}$$

To produce the simplex tableau corresponding to the new *b.f.s.*, we just have to arrange that the columns of $\mathbf{I}_m$ correspond to the new basic variables, and that in the new *e.c.c.s*, c^*_t, corresponding to x_t, is zero. Having chosen s and t, all this is achieved by the following operations:

① Divide the s-th row of the tableau by a'_{st},

i.e. $a^*_{sj} = a'_{sj}/a'_{st}$, $j = 1, 2, \ldots, n$,

$$b^*_s = b'_s/a'_{st} \tag{7}$$

(note that $a'_{st} > 0$, so $b^*_s \geq 0$).

② For $i = 1, 2, \ldots, s-1, s+1, \ldots, m$ add $-a'_{it} \times$ (new s-th row) to i-th row,

i.e. $a^*_{ij} = a'_{ij} - a'_{it} \times a^*_{sj}$, $j = 1,2,\ldots,n$, $i = 1,2,\ldots,m$, $i \neq s$

$$= a'_{ij} - a'_{it} \times \frac{a'_{sj}}{a'_{st}}; \text{ and } b^*_i = b'_i - a'_{it} \times \frac{b'_s}{a'_{st}}. \tag{8}$$

③ Add $-c'_t \times$ (new s-th row) to the c-row,

$$\text{i.e.} \quad c^*_j = c'_j - c'_t \times \frac{a'_{sj}}{a'_{st}}, \; j = 1, \ldots, n, \text{ and } \alpha^* = \alpha' - c'_t \frac{b'_s}{a'_{st}}. \tag{9}$$

We shall refer to these as operations ①, ②, and ③ of the simplex method. Referring back to section 3.1 we see that:

① effectively expresses the new basic variable x_t in terms of the new set of non-basic variables, namely $x_s, x_{m+1}, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n$,

② eliminates x_t from the i-th equality constraint and introduces x_s ($i \neq s$) so that the i-th equation now expresses the basic variable x_i in terms of the new non-basic variables,

③ eliminates x_t from the expression for f and produces a linear combination of the new set of non-basic variables.

The new tableau has the following form:

$$
\begin{array}{|cccccccc|cccccc|c|}
\hline
1 & 0 & \dots & 0 & a^*_{1s} & 0 & \dots & 0 & a^*_{1\,m+1} & a^*_{1\,m+2} & \dots & 0 & \dots & a^*_{1n} & b^*_1 \\
0 & 1 & \dots & 0 & a^*_{2s} & 0 & \dots & 0 & a^*_{2\,m+1} & a^*_{2\,m+2} & \dots & 0 & \dots & a^*_{2n} & b^*_2 \\
\vdots & & & & & & & & & & & & & & \\
0 & 0 & \dots & 0 & a^*_{ss} & 0 & \dots & 0 & a^*_{s\,m+1} & a^*_{s\,m+2} & \dots & 1 & \dots & a^*_{sn} & b^*_s \\
\vdots & & & & & & & & & & & & & & \\
0 & 0 & \dots & . & a^*_{ms} & 0 & \dots & 1 & a^*_{m\,m+1} & a^*_{m\,m+2} & \dots & 0 & \dots & a^*_{mn} & b^*_m \\
\hline
0 & 0 & \dots & 0 & c^*_s & 0 & \dots & 0 & c^*_{m+1} & c^*_{m+2} & \dots & 0 & \dots & c^*_n & f+\alpha^* \\
\hline
\end{array}
$$

and we may repeat the procedure until either all *e.c.c.s* are non-negative or we find a negative *e.c.c.* underneath a column of coefficients which are all non-positive.

3.3 To Summarise

The criterion for optimality is $c'_j \geqslant 0$, $j = 1, 2, \dots, n$; the criterion for feasible solutions unbounded below is $c'_t < 0$ and $a'_{it} \leq 0$, $i = 1, 2, \dots, m$, for any non-basic t.

It is worth looking back at the simplex procedure described in the previous section to check that it is no more difficult to carry out when the columns of $\mathbf{I}_m$ are in random columns of the tableau. The assumption that the first (or last) m columns are $\mathbf{I}_m$ is a very convenient one for descriptive and theoretical purposes, so we shall make this assumption frequently, but in practice, of course, one has to perform the appropriate operations whatever the positions of the columns of $\mathbf{I}_m$, i.e. whichever variables are basic.

The choice of *min* c'_j to identify t is not essential. Any negative *e.c.c.* can be used, but this choice is commonly used in practice.

We shall refer to the element a'_{st} as the *pivot*.

The integers s and t identify the basic variable that is to become non-basic and the non-basic variable that is to take its place. In terms of the feasible region R one stage of the simplex method consists of moving from one extreme point along an edge of R to an adjacent extreme point where f has a lower value (provided that $b'_s > 0$).

3.4 Example

Maximise $x_1 + 2x_2 + x_3$ *subject to*

$$
\begin{aligned}
2x_1 + x_2 - x_3 &\leq 2, \\
2x_1 - x_2 + 5x_3 &\leq 6, \\
4x_1 + x_2 + x_3 &\leq 6, \quad \textit{and} \quad x_1, x_2, x_3 \geq 0.
\end{aligned}
$$

In canonical form we have *minimise* $(-1,-2,-1,0,0,0)\mathbf{x}$ *subject to*

$$\begin{pmatrix} 2 & 1 & -1 & 1 & 0 & 0 \\ 2 & -1 & 5 & 0 & 1 & 0 \\ 4 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 6 \\ 6 \end{pmatrix}.$$

Here $\mathbf{A} \supset \mathbf{I}_3$, and cost coefficients corresponding to basic variables x_4, x_5, x_6 are zero, so we have the initial simplex tableau

$$\begin{array}{cccccc|c|l} x_1 & x_2 & x_3 & \underline{x_4} & \underline{x_5} & \underline{x_6} & & \theta \\ 2 & ① & -1 & 1 & 0 & 0 & 2 & 2 \leftarrow \\ 2 & -1 & 5 & 0 & 1 & 0 & 6 & \\ 4 & 1 & 1 & 0 & 0 & 1 & 6 & 6 \\ \hline -1 & -2 & -1 & 0 & 0 & 0 & f+0 & \\ & \uparrow & & & & & & \end{array}$$

Here we have written the variable names above the columns and underlined the basic variables. This helps at first but should soon be unnecessary, whereas a consistent notation to define the steps of the method is strongly recommended.

Thus $\uparrow$ indicates the *min* c_j', the *θ-column* lists the ratios b_i'/a_{it}' for $a_{it}' > 0$ and $\leftarrow$ indicates the minimum of these values. The circle indicates the pivot.

In this tableau $t = 2$, $s = 1$, x_4 is to become non-basic and x_2 is to become basic.

Now we carry out the computational operations ①, ②, and ③ and continue.

$$\begin{array}{cccccc|c|l} 2 & 1 & -1 & 1 & 0 & 0 & 2 & \\ 4 & 0 & ④ & 1 & 1 & 0 & 8 & 2 \leftarrow \\ 2 & 0 & 2 & -1 & 0 & 1 & 4 & 2 \\ \hline 3 & 0 & -3 & 2 & 0 & 0 & f+4 & \\ & & \uparrow & & & & & \end{array}$$

$$\begin{array}{cccccc|c|} 3 & 1 & 0 & \frac{5}{4} & \frac{1}{4} & 0 & 4 \\ 1 & 0 & 1 & \frac{1}{4} & \frac{1}{4} & 0 & 2 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & 0 \\ \hline 6 & 0 & 0 & \frac{11}{4} & \frac{3}{4} & 0 & f+10 \end{array}$$

In the third tableau all elements in the c-row are ≥ 0 so we have the optimum solution which, from the tableau, is $x_1 = x_4 = x_5 = 0$ (the non-basic variables) and $x_2 = 4$, $x_3 = 2$, $x_6 = 0$ (the basic variables).

The optimum value is given by $f + 10 = 0$, i.e. $f_{opt} = -10$; remember that the c-row is a shorthand for

$$6x_1 + 0x_2 + 0x_3 + \tfrac{11}{4}x_4 + \tfrac{3}{4}x_5 + 0x_6 = f + 10.$$

We notice in this case that the optimum solution, $\mathbf{x} = (0,4,2,0,0,0)^T$, is degenerate, and also that in the second tableau there was a tie in the θ-column. This tie indicates that the next *b.f.s.* will be degenerate, but we observe that this does not cause the simplex procedure to break down. We can also verify that if, in the second tableau, we choose $s = 3$ we get the *same solution* provided by the alternative third tableau, but the tableau is different (*ER*). We can also verify that a different choice of t in the first tableau also leads to the same solution (*ER*).

3.5

The crucial question now is whether or not we can be sure of obtaining the optimum solution in a finite number of simplex stages.

Theorem 5

If all *b.f.s.s* of a *l.p.p.* are non-degenerate the simplex method terminates in a finite number of stages■

The proof of this important result is very simple.

The feasible region R has a finite number of extreme points. If the simplex method is not finite then at some stage it must produce an extreme point (*b.f.s.*) already encountered. However, at any point, and in particular at each extreme point, the objective function has a unique value and this value decreases at each stage by the positive amount

$$-c_t' \times b_s'/a_{st}'.$$

Thus the objective function values produced by the simplex method are strictly monotone decreasing and so any particular extreme point can only appear once as the current *b.f.s.*■

If a *l.p.p.* has degenerate *b.f.s.s* then the simplex method could possibly *cycle* round a sequence of identical degenerate extreme points because then we could have $b_s' = 0$ and hence $\theta = 0$. However, this need not happen as the examples in section 3.4 and exercise 4.1 show. We return to the problem of cycling in chapter 4 where we also deal with the two remaining major problems $\mathbf{A} \not\supset \mathbf{I}_m$ and $r(\mathbf{A}) < m$, neither of which requires the method we have already described to be changed.

3.6

We can verify that there is a simplex tableau for any *b.f.s.* and in doing so we see the relationship between the original coefficients $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$, and the coefficients of the tableau $\mathbf{A}'$, $\mathbf{b}'$, $\mathbf{c}'$.

Let $\mathbf{x}$ be any *b.f.s.* and *w.l.o.g.* assume that $x_1, x_2, \ldots, x_m$ are the basic variables. Partition $\mathbf{x}$ into $\mathbf{x}_1 = (x_1, x_2, \ldots, x_m)^T$ and $\mathbf{x}_2 = (x_{m+1}, x_{m+2}, \ldots, x_n)^T$, $\mathbf{c}$ similarly, and $\mathbf{A}$ into $(\mathbf{A}_1, \mathbf{A}_2)$, where $\mathbf{A}_1$ is the first m columns of $\mathbf{A}$, and $\mathbf{A}_1$ is non-singular.

Then $\mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = \mathbf{b}$, so

$$\mathbf{x}_1 = \mathbf{A}_1^{-1}\mathbf{b} - \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{x}_2 \geq \mathbf{0},$$

and with $\mathbf{x}_2 = \mathbf{0}$, $\mathbf{x} = \begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}$ is the *b.f.s.*

Referring back to the first tableau in section 3.2,

$$a'_{ij} = (\mathbf{A}_1^{-1}\mathbf{A})_{ij} \quad \text{and} \quad b'_i = (\mathbf{A}_1^{-1}\mathbf{b})_i.$$

Alternatively $\mathbf{a}'_{*j}$ is the j-th column of the matrix $\mathbf{A}_1^{-1}\mathbf{A}$.

The objective function

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{c}^T\mathbf{x} = \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2 \\ &= \mathbf{c}_1^T\mathbf{A}_1^{-1}\mathbf{b} + (-\mathbf{c}_1^T\mathbf{A}_1^{-1}\mathbf{A}_2 + \mathbf{c}_2^T)\mathbf{x}_2 \\ &= -\alpha' + c'_{m+1}x_{m+1} + \ldots + c'_n x_n = -\alpha' + \mathbf{c}_2'^T\mathbf{x}_2. \end{aligned}$$

So this *b.f.s.* is optimal if the vector of non-basic *e.c.c.s* is non-negative, i.e. $-\mathbf{c}_1^T\mathbf{A}_1^{-1}\mathbf{A}_2 + \mathbf{c}_2 \geq \mathbf{0}$.

This analysis does not give us an alternative method for solving *l.p.p.s* because in general we do not know which m variables will be basic in the optimum solution, and so we do not know which m columns of $\mathbf{A}$ to choose as $\mathbf{A}_1$. It does tell us how each successive system of constraint equations and *e.c.c.s* is related to the original system, and we may notice that if $\mathbf{A} \supset \mathbf{I}_m$ then at any stage the appropriate $\mathbf{A}_1^{-1}$ will actually be present in the simplex tableau, occupying the columns of the tableau which were originally the columns of $\mathbf{I}_m$.

For many people the simplex method in particular the tableau, has a mysterious air about it. It is true that it is still not completely clear why the simplex method works as well as it does in practice, but the reasons why the method works at all and the purpose of the operations at each stage present no difficulties if we remember that each row of the tableau represents an equation (see (1) and (2) of section 3.2) and if we bear in mind the interpretation of steps ①, ② and ③ of the simplex method given towards the ends of sections 3.1 and 3.2.

3.7

There is another useful way of representing one stage of the simplex method. Operations ① and ② consist of elementary row operations on the augmented matrix $(\mathbf{A}',\mathbf{b}')$. Performing an elementary row operation on any $m \times n$ matrix $\mathbf{A}$ is equivalent to performing the operation on the unit matrix $\mathbf{I}_m$, to obtain a matrix $\mathbf{E}$ say, and then pre-multiplying $\mathbf{A}$ by $\mathbf{E}$. The matrix $\mathbf{E}$ is called an *elementary matrix*.

For example $\mathbf{E}_s\mathbf{A}$, where

$$\mathbf{E}_s = \begin{pmatrix} 1 & 0 & . & . & . & . & . & . & 0 \\ 0 & 1 & . & . & . & . & . & . & 0 \\ & & . & & & & & & \\ & & & 1 & & & & & \\ & & & & \lambda & & & & \\ & & & & & 1 & & & \\ & & & & & & . & & \\ & & & & & & & . & \\ 0 & 0 & . & . & . & . & . & . & 1 \end{pmatrix} \quad \text{and} \quad (\mathbf{E})_{ss} = \lambda,$$

is the matrix obtained by multiplying the s-th row of $\mathbf{A}$ by λ, and the matrix $\mathbf{E}_i\mathbf{A}$, where

$$\mathbf{E}_i = \begin{pmatrix} 1 & 0 & . & . & . & . & . & . & 0 \\ 0 & 1 & . & . & . & . & . & . & 0 \\ . & & & & & & & & \\ . & & & 1 & & & & & \\ 0 & & & & & 1 & . & . & 0 \\ . & & & & & . & & & \\ . & & & \lambda & & & . & & \\ . & & & & & & & . & \\ 0 & . & . & . & . & . & . & . & 1 \end{pmatrix} \quad \text{and} \quad (\mathbf{E}_i)_{is} = \lambda,$$

is the matrix obtained by adding $\lambda \times$ (s-th row of $\mathbf{A}$) to the i-th row of $\mathbf{A}$.

Thus, referring to section 3.2,

$(\mathbf{A}^*,\mathbf{b}^*) = \mathbf{E}_m \ldots \mathbf{E}_{s+1}\mathbf{E}_{s-1} \ldots \mathbf{E}_2\mathbf{E}_1\mathbf{E}_s(\mathbf{A}',\mathbf{b}')$,

where λ in $\mathbf{E}_s$ is $1/a'_{st}$, and

λ in $\mathbf{E}_i$ is $-a'_{it}$, $i = 1,2, \ldots, s-1, s+1, \ldots, m$.

The special form of $\mathbf{E}_i$ results in a simple form for the matrix product $\mathbf{E}_m\mathbf{E}_{m-1} \ldots \mathbf{E}_{s+1}\mathbf{E}_{s-1} \ldots \mathbf{E}_2\mathbf{E}_1$, namely

$$\mathbf{E}_k^* = \begin{pmatrix} 1 & \cdots & 0 & -a'_{1t}/a'_{st} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & 1 & -a'_{s-1,t}/a'_{st} & 0 & & \vdots \\ \vdots & & 0 & 1/a'_{st} & 0 & & \vdots \\ \vdots & & 0 & -a'_{s+1,t}/a'_{st} & 1 & & \vdots \\ \vdots & & 0 & -a'_{s+2,t}/a'_{st} & 0 & \ddots & \vdots \\ \vdots & & \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & -a'_{mt}/a'_{st} & 0 & \cdots & 1 \end{pmatrix},$$

where k denotes the number of the simplex stage (ER).

In the example in section 3.4

$$\mathbf{E}_1^* = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad \mathbf{E}_2^* = \begin{pmatrix} 1 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & -\frac{2}{4} & 1 \end{pmatrix}$$

and the product $\mathbf{E}_2^* \mathbf{E}_1^*$ is the matrix $\mathbf{A}_1^{-1}$ of section 3.4 for the final tableau.

We note that the columns 1, 2 and 3 of $\mathbf{A}_1^{-1}$,

$$\mathbf{A}_1^{-1} = \begin{pmatrix} \frac{5}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 \\ -\frac{3}{2} & -\frac{1}{2} & 1 \end{pmatrix} = \mathbf{E}_2^* \mathbf{E}_1^*,$$

appear in columns 4, 5, and 6 of the final tableau because columns 4, 5, and 6 of the original matrix $\mathbf{A}$ were columns 1, 2, and 3 of $\mathbf{I}_3$.

For the *e.c.c.s* $\mathbf{c}^*$ we note that if $\mathbf{e}_s^T$ is the m-vector $(0,0,\ldots 0,1,0 \ldots 0)$, where $(\mathbf{e}_s)_s = 1$, then the n-vector $\mathbf{e}_s^T \mathbf{A}'$ is the s-th row of $\mathbf{A}'$.

Hence

$$\mathbf{c}^{*T} = \mathbf{c}'^T + (-c'_t/a'_{st})\mathbf{e}_s^T \mathbf{A}',$$

and as $(\frac{1}{a'_{st}})\mathbf{e}_s^T \mathbf{A}'$ is just the s-th row of $\mathbf{A}^*$, we have

$$\mathbf{c}^{*T} = \mathbf{c}'^T + (-c'_t \mathbf{e}_s^T)\mathbf{A}^*.$$

Comparing the first and second tableaux of the example in section 3.4, $t = 2$, $c'_t = -2$, $s = 1$, $a'_{st} = 1$, so $\mathbf{e}_s^T = (1,0,0)$ and

$$\mathbf{e}_s^T\mathbf{A}^* = (2,1,-1,1,0,0) \quad (= \mathbf{e}_s^T\mathbf{A} \text{ in this case}).$$

Thus the *e.c.c.s* in the second tableau, (3,0,–3,2,0,0), are (–1,–2,–1, 0,0,0) + 2(2,1,–1,1,0,0).

We note in passing that our observations about row operations and elementary matrices apply in exactly the same way to column operations if we replace *pre-multiplication* by *post-multiplication*. Thus to perform an elementary column operation on **A**, we perform the operation on $\mathbf{I}_n$ to produce the elementary matrix **E**, and then form the product **AE**.

To illustrate this point, we consider the remaining elementary matrix operation, that of interchanging two rows (or columns).

The matrix

$$\mathbf{E} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ is } \mathbf{I}_3 \text{ with the first and second rows}$$

(or columns) interchanged, and for

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$\mathbf{EA} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \text{and} \quad \mathbf{AE} = \begin{pmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{pmatrix}.$$

The product of several elementary matrices of this type is called a permutation matrix **P**. Each element of **P** is 0 or 1 and there is precisely a single 1 in each row or column of **P**. Pre-multiplying a matrix **A** by **P** re-orders or permutes the rows of **A**; post-multiplying **A** by **P** permutes the columns of **A**.

The simplex method was devised by G. Dantzig in 1947. The development of the simplex method in chapters 1, 2, 3 is more or less standard, and most texts follow this treatment. The basically similar developments available in {9}, {10} and {12} are useful further reading.

Exercises 3

1. Solve the *l.p.p.s* in canonical form defined by the following data:

(i) $$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & -2 \\ 0 & 1 & 0 & 2 & -3 & 1 \\ 0 & 0 & 1 & 2 & -5 & 6 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 5 \\ 3 \\ 5 \end{pmatrix},$$

$\mathbf{c}^T = (0,0,0,-3,8,-5)$.

(ii) $$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 6 \\ 3 & 1 & -4 & 0 & 0 & 2 \\ 1 & 0 & 2 & 0 & 1 & 2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 9 \\ 2 \\ 6 \end{pmatrix},$$

$\mathbf{c}^T = (5,0,-9,0,0,-5)$.

2. In the simplex method the decrease in $f(\mathbf{x})$ at any stage is not necessarily the largest possible. Explain why, and explain what additional calculations are needed to achieve the largest possible decrease in $f(\mathbf{x})$ at each simplex stage. (In practice the total number of stages needed is not reduced enough in general to make the extra calculations worthwhile.)

3. Consider the *l.p.p.*

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{Ax} = \mathbf{b}, \ x_{k+1}, \ldots, x_n \geq 0,$$

i.e. there are k free variables $x_1, x_2, \ldots, x_k$. Explain how a *l.p.p.* in canonical form of size $(m-k) \times (n-k)$ could be obtained instead of one of size $m \times (n+k)$ that would result from the conversion technique described in section 2.1 (iv).

4. In a certain *l.p.p.* x_i is a free variable and is replaced by $u_j - v_j$, $u_j, v_j \geq 0$, to obtain canonical form. Can u_j and v_j both be basic variables in the same *b.f.s.*?

5. In the simplex method explain why the optimality test, $c'_j \geq 0$, $j = 1, 2, \ldots, n$, is a sufficient but not a necessary test. (Hint: consider degeneracy.)

6. Suppose that in a final (optimum) simplex tableau $c'_j = 0$ for some j where x_j is non-basic. Explain how a different optimum solution could usually be obtained, and explain why only usually. Hence state a simple sufficient criterion for deciding whether or not an optimum solution is unique, explain how all basic optimum solutions can be found and define, constructively, the set of all optimum solutions.

7. What modifications to the simplex method as described are needed to solve *directly* problems of the form

$$\textit{maximise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}?$$

Try them out on 1(i) above with $\mathbf{c}^T = (0,0,0,3,-8,5)$.

8. A *l.p.p.* with a finite optimum solution is being solved by the simplex method and in one tableau, which is *not* optimum, a single basic variable is zero. Prove that this tableau cannot reappear at a later stage.

9. For the examples 1(i) and 1(ii) above, write down the matrices $\mathbf{E}_k^*$. If the augmented matrix $(\mathbf{A}',\mathbf{b}')$ of the optimum tableau is $\mathbf{A}_1^{-1}(\mathbf{A},\mathbf{b})$, identify $\mathbf{A}_1^{-1}$ in the optimum tableau and verify that

$$(\mathbf{A}',\mathbf{b}')_{opt} = \mathbf{A}_1^{-1}(\mathbf{A},\mathbf{b})_{orig}$$

and that $\mathbf{A}_1^{-1}$ is the product of the matrices $\mathbf{E}_k^*$.

10. At any stage of the simplex method the columns of $\mathbf{A}$ corresponding to the current basic variables are linearly independent. If the j-th column of $\mathbf{A}$, $\mathbf{a}_{*j}$, is expressed as a linear combination of them, show that the coefficients of the linear combination are the elements of $\mathbf{a}'_{*j}$, the j-th column of $\mathbf{A}'$.

NOTES

NOTES

CHAPTER 4

THE SIMPLEX METHOD CONTINUED

4.1

The simplex method does not provide a guaranteed method for solving *l.p.p.s* in general until the three problems mentioned at the end of section 3.5 are dealt with. We will refer to them as *the initial tableau, rank deficiency* and *cycling* and discuss them in that order. Before doing so, some preliminary comments are worthwhile. We should expect that in general, after conversion to canonical form, $\mathbf{A}$ will have rank m and $\mathbf{I}_m$ will not be present. If $r(\mathbf{A}) < m$, then we cannot, by row operations on $\mathbf{A}$, produce $\mathbf{I}_m$ among the columns of $\mathbf{A}$, since row operations do not change the rank. Again, if $r(\mathbf{A}) < m$, then there are not m independent columns of $\mathbf{A}$ and so there are no basic solutions with m basic variables. This does not mean that R is empty, just that basic solutions will only have $k = r(\mathbf{A})$ basic variables. Rather than modify the simplex procedure to take account of this, it is preferable to remove linearly dependent rows and then continue with a matrix $\mathbf{A}$ that has full rank. However, if we have rank deficiency then we cannot have $\mathbf{I}_m$ contained in $\mathbf{A}$ so it is natural to deal with obtaining the initial tableau first. Note that if $r(\mathbf{A}) = m$ as expected, we cannot simply reduce m chosen (independent) columns to $\mathbf{I}_m$ by row operations since we have no way of knowing that the resulting vector $\mathbf{b}$ will be non-negative. It is possible that a *b.f.s.* is known even when $\mathbf{A} \not\supset \mathbf{I}_m$, but as this situation still requires almost as much work to produce the initial simplex tableau as the method we describe below, we ignore this possibility.

Obtaining the initial tableau and dealing with rank deficiency requires consideration of a sequence of possible cases, and it is important to remember that we need a definite, algorithmic procedure.

4.2

We first mention a trivial, but very important, situation.

Suppose, with the problem in canonical form, $\mathbf{b} \geq \mathbf{0}$ and $\mathbf{A} \supset \mathbf{I}_m$, the cost coefficients corresponding to the basic variables are non-zero.

Then, although $(\mathbf{A},\mathbf{b})$ has the required form, $\mathbf{c}$ does not and we do not yet have a simplex tableau as defined in section 3.2.

To produce an appropriate set of *e.c.c.s* we add $-c_{j_i} \times$ (i-th row of $\mathbf{A}$) to $\mathbf{c}^T$, for $i = 1, 2, \dots, m$, where the j_i-th column of $\mathbf{A}$ is the i-th column of $\mathbf{I}_m$, $i = 1, 2, \dots, m$.

For example, in exercise 3.1(ii) where

$$(\mathbf{A},\mathbf{b}) = \left(\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 6 & 9 \\ 3 & 1 & -4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 2 & 0 & 1 & 2 & 6 \end{array}\right)$$

suppose the objective function were

$$f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = (1,-2,1,1,1,-1)\mathbf{x}, \tag{1}$$

$$\text{then}\quad j_1 = 4,\ j_2 = 2,\ j_3 = 5 \quad \text{and} \tag{2}$$

$$\begin{aligned} \mathbf{c}'^T = \mathbf{c}^T &- 1(1,0,\ 0,1,0,6) \\ &+ 2(3,1,-4,0,0,2) \\ &- 1(1,0,\ 2,0,1,2) \\ &= (5,0,-9,0,0,-5). \end{aligned} \tag{3}$$

These are the objective function cost coefficients given in exercise 3.1(ii), but notice that these row operations on $\mathbf{A}$ must also be applied to the $\begin{pmatrix}\mathbf{b}\\ \alpha\end{pmatrix}$ column, so instead of $f + 0$ in the bottom right-hand corner, we would start the simplex procedure with

$$f + 0 + (-1)\times 9 + 2\times 2 + (-1)\times 6 = f - 11. \tag{4}$$

This particular part of the conversion to simplex tableau form is easily done using the tableau and can be indicated as follows:

$$\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 6 & 9 \\ 3 & 1 & -4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 2 & 0 & 1 & 2 & 6 \\ \hline 1 & -2 & 1 & 1 & 1 & -1 & 0 \\ 5 & 0 & -9 & 0 & 0 & -5 & -11 \end{array} \tag{5}$$

If, for convenience, we assume that $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$, $\mathbf{A}_1 = \mathbf{I}_m$, $\mathbf{c}^T = (\mathbf{c}_1^T, \mathbf{c}_2^T)$ and $\mathbf{c}_1^T \neq \mathbf{0}_m^T$, then

$$\mathbf{c}'^T = \mathbf{c}^T - \mathbf{c}_1^T\mathbf{A} = (\mathbf{c}_1'^T, \mathbf{c}_2'^T), \quad \text{where} \quad \mathbf{c}_1'^T = \mathbf{0}_m. \tag{6}$$

4.3 Obtaining the Initial *b.f.s.* and Simplex Tableau when $\mathbf{A} \not\supset \mathbf{I}_m$

We assume the *l.p.p.* is in canonical form and for convenience in this section we assume $r(\mathbf{A}) = m$. We also assume that no columns of $\mathbf{I}_m$ are present among the columns of $\mathbf{A}$, leaving the intermediate case, when some but not all are present, to exercise 4.2.

Consider the two *l.p.p.s:*

I *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}_n$, and

II *minimise* $(\mathbf{0}^T, \mathbf{e}^T)\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix}$ *subject to* $(\mathbf{A}, \mathbf{I}_m)\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} = \mathbf{b}$, $\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} \geq \mathbf{0}_{n+m}$, where $\mathbf{e}$ is the m-vector $(1, 1, \dots, 1)^T$.

Problem II is in canonical form and, if written as

$$\textit{minimise } \tilde{\mathbf{c}}^T\tilde{\mathbf{x}} \textit{ subject to } \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}},\ \tilde{\mathbf{x}} \geq \mathbf{0},$$

we see that $\tilde{\mathbf{A}} \supset \mathbf{I}_m$ and, using section 4.2, we easily obtain the initial simplex tableau and hence the optimum solution of problem II.

The variables $z_1, \dots, z_m$ are called *artificial variables,* and problem II is concerned with minimising their sum. Notice that $\tilde{f} = \tilde{\mathbf{c}}^T\tilde{\mathbf{x}} \geq 0$, and that problem II is feasible because $\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{b}\end{pmatrix}$ is the *b.f.s.* of the initial tableau for problem II, so problem II has a finite optimum value $\tilde{f}_{opt}$, and $\tilde{f}_{opt} \geq 0$. (Remember $\mathbf{b} \geq \mathbf{0}$, see section 2.2 and the end of section 2.9.)

(i) *If* $\tilde{f}_{opt} > 0$ *then problem I has no feasible solutions.* For suppose $\mathbf{x}_0$ is a feasible solution for problem I, then $\tilde{\mathbf{x}}_0 = \begin{pmatrix}\mathbf{x}_0\\ \mathbf{0}\end{pmatrix}$ is a feasible solution for problem II with value zero. By the same argument, if problem II has any feasible solutions with value zero, then problem I has a feasible solution.

(ii) *If* $\tilde{f}_{opt} = 0$ *then problem* I *is feasible,* but we do not necessarily have a *b.f.s.* of problem I. There are two possibilities: either all the basic variables in the optimum solution of II are among $x_1, x_2, \dots, x_n$, or some basic variables are artificial variables with value zero. Remember that whatever set of *e.c.c.s* we produce when solving problem II we are still minimising $\Sigma_{i=1}^m z_i$, so if $\tilde{f}_{opt} = 0$ then $\mathbf{z} = \mathbf{0}$ in the optimum solution.

Leaving the second possibility to (iii) below, we consider the first possibility, which we expect to be the case in general. Since all the z_i are non-basic, we can assume *w.l.o.g.* that the basic variables are $x_1; x_2, \dots, x_m$ and that $\mathbf{I}_m$ is in the first m columns of the optimum tableau. This tableau has the following form:

$x_1 \quad \dots \quad x_m$	$x_{m+1} \quad \dots \quad x_n$	$z_1 \quad \dots \quad z_n$	
$\mathbf{I}_m$	$\mathbf{A}'_2$	$\mathbf{A}'_3$	$\mathbf{b}'$
$0 \quad \dots \quad 0$	$\tilde{c}'_{m+1} \quad \dots \quad \tilde{c}'_n$	$\tilde{c}'_{n+1} \quad \dots \quad \tilde{c}'_{n+m}$	0

(1)

and to obtain an initial simplex tableau for the original problem, problem I, we simply remove the columns corresponding to $z_1, \ldots, z_m$, replace the coefficients $0, \ldots, 0, \tilde{c}'_{m+1}, \ldots, \tilde{c}'_n$ in the c-row by the original cost coefficients $\mathbf{c}^T$, and convert this new c-row to the form $0, \ldots, 0, c'_{m+1}, \ldots, c'_n$ as described in section 4.2.

It is convenient to call problem II the *artificial problem,* and solving it, *part* I. Solving the original problem, starting with the initial simplex tableau with coefficients $\mathbf{A}'$, $\mathbf{b}'$, $\mathbf{c}'$, is called *part* II. Together they constitute the *two-part simplex method* for solving *any l.p.p.* (in canonical form). (The name *two-stage simplex method* is sometimes used, but we shall use *'stage'* to refer to one of the steps of the simplex method as described in section 3.2.) A simple example to illustrate the two-part method is given in section 4.4 below.

Notice that if the matrix $\mathbf{A}$ of the original problem (problem I) is partitioned

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2), \text{ where } \mathbf{A}_1 \text{ is } m \times m,$$

then $\mathbf{A}'_3$ in (1) must be $\mathbf{A}_1^{-1}$, $\mathbf{b}' = \mathbf{A}_1^{-1}\mathbf{b}$ and $\mathbf{A}'_2 = \mathbf{A}_1^{-1}\mathbf{A}_2$.

It is worth recalling the remarks near the ends of the sections 2.9 and 4.1. Just as, to solve a *l.p.p.* we need to solve a particular $m \times m$ system of equations and the simplex method is in effect an efficient way of finding that system, so, to find an initial simplex tableau we need to reduce m columns of $\mathbf{A}$ to $\mathbf{I}_m$ by row operations and the artificial problem provides an efficient way of choosing an appropriate set of m columns.

(iii) If $\tilde{f}_{opt} = 0$, but the basic variables in $\tilde{\mathbf{x}}_{opt}$ include some artificial variables, then, supposing *w.l.o.g.* that the basic variables are $x_1, x_2, \ldots, x_k, z_{k+1}, z_{k+2}, \ldots, z_m$, the final tableau for problem II has the form below, where the c-row, which is to be discarded, has been omitted:

$$\begin{array}{c|c|c|c|c|}
\multicolumn{1}{c}{x_1 \;\cdots\; x_k} & \multicolumn{1}{c}{x_{k+1} \;\cdots\; x_n} & \multicolumn{1}{c}{z_1 \;\cdots\; z_k} & \multicolumn{1}{c}{z_{k+1} \;\cdots\; z_m} & \multicolumn{1}{c}{} \\
\hline
\begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} & \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{O} & \begin{matrix} b'_1 \\ \vdots \\ b'_k \end{matrix} \\
\hline
\mathbf{O} & \mathbf{B}_3 & \mathbf{B}_4 & \begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\
\hline
\end{array} \qquad (2)$$

The matrix $\mathbf{B}_1$ is $k \times (n-k)$, $\mathbf{B}_2$ is $k \times k$, $\mathbf{B}_3$ is $(m-k) \times (n-k)$ and $\mathbf{B}_4$ is $(m-k) \times k$.

To obtain the initial tableau for problem I we just perform appropriate row operations to produce the *missing* columns of $\mathbf{I}_m$ in the $\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_3 \end{pmatrix}$ part of the tableau. These operations are just operations ① and ② of the simplex method, using *any* non-zero element of $\mathbf{B}_3$ as pivot. For example, suppose $(\mathbf{B}_3)_{11} \neq 0$, $((\mathbf{B}_3)_{11} = a'_{k+1,k+1})$. Then operations ① and ② will produce $\mathbf{e}_{k+1}$ in the $(k+1)$-th column (and remove $\mathbf{e}_{k+1}$ from the $(k+1)$-th artificial column), x_{k+1} will become basic and z_{k+1} non-basic. This effectively reduces the size of $\mathbf{B}_3$ by one row and one column, and we continue until $\mathbf{B}_3$ disappears completely, at which time the m columns of $\mathbf{I}_m$ are present in the first n columns of the tableau and we have the situation discussed in (ii) above. A non-zero element can always be found in $\mathbf{B}_3$ (ER). Notice that the *b-column* is unchanged because $b'_{k+1} = b'_{k+2} = \ldots = b'_m = 0$. The m artificial columns may be removed. There are sometimes reasons for leaving them, in which case they provide a record, as in (ii) above, of the operations which have transformed the system from $(\mathbf{A},\mathbf{b})$ to $(\mathbf{A}',\mathbf{b}')$, where $\mathbf{A}' \supset \mathbf{I}_m$.

4.4 Example

Solve the *l.p.p.*

minimise $x_1 + 5x_2 + 2x_3 + 2x_4 + 7x_5$ *subject to*

$$\begin{pmatrix} 1 & 1 & 1 & -1 & 0 \\ 1 & 2 & 1 & 0 & -1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \ \mathbf{x} \geq 0.$$

No columns of $\mathbf{I}_2$ are present among the columns of $\mathbf{A}$ so we use the two-part method, with artificial variables z_1 and z_2. The artificial problem is *minimise* $z_1 + z_2 = \tilde{f}(\tilde{\mathbf{x}}) = (0,0,0,0,0,1,1)\begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix}$ *subject to*

$$\begin{pmatrix} 1 & 1 & 1 & -1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \ \begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} \geq \mathbf{0}.$$

The initial tableau (for the artificial problem) and the simplex calculations now follow:

$$\begin{array}{ccccc|cc|c|cl}
1 & 1 & 1 & -1 & 0 & 1 & 0 & 2 & 2 \leftarrow & \\
1 & 2 & 1 & 0 & -1 & 0 & 1 & 5 & \frac{5}{2} & (1) \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 1 & \tilde{f} & & \\
-2 & -3 & -2 & 1 & 1 & 0 & 0 & \tilde{f}-7 & & \\
 & \uparrow & & & & & & & &
\end{array}$$

$$\begin{array}{ccccc|cc|c|cl}
1 & 1 & 1 & -1 & 0 & 1 & 0 & 2 & & \\
-1 & 0 & -1 & 2 & -1 & -2 & 1 & 1 & \frac{1}{2} \leftarrow & (2) \\
\hline
1 & 0 & 1 & -2 & 1 & 3 & 0 & \tilde{f}-1 & & \\
 & & & \uparrow & & & & & &
\end{array}$$

$$\begin{array}{ccccc|cc|c|l}
\frac{1}{2} & 1 & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & \frac{1}{2} & \frac{5}{2} & \\
-\frac{1}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & -1 & \frac{1}{2} & \frac{1}{2} & (3) \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 1 & \tilde{f} &
\end{array}$$

At this stage $\tilde{\mathbf{c}}' \geq \mathbf{0}$, so we have $\tilde{f}_{opt}$; $\tilde{f}_{opt} = 0$ so the feasible region R for the original problem is not empty.

The initial *b.f.s.* for the original problem is provided by the tableau (3) and is

$$x_2 = \tfrac{5}{2},\ x_4 = \tfrac{1}{2},\ x_1 = x_3 = x_5 = 0$$

and the equivalent system of equations $\mathbf{A}'\mathbf{x} = \mathbf{b}'$, with $\mathbf{I}_2 \subset \mathbf{A}'$ is

$$\mathbf{A}' = \begin{pmatrix} \frac{1}{2} & 1 & \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \end{pmatrix}, \quad \mathbf{b}' = \begin{pmatrix} \frac{5}{2} \\ \frac{1}{2} \end{pmatrix} . \tag{4}$$

To obtain the initial simplex tableau for the original problem, i.e. the initial tableau of part II, we remove the artificial columns, replace the c-row by (1,5,2,2,7) and convert the cost coefficients to *e.c.c.s* with $c_2' = c_4' = 0$.

$$\begin{array}{ccccc|ccl}
\frac{1}{2} & 1 & \frac{1}{2} & 0 & -\frac{1}{2} & \frac{5}{2} & 5 \leftarrow & \\
-\frac{1}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \frac{1}{2} & & (5) \\
\hline
1 & 5 & 2 & 2 & 7 & 0 & & \\
-\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{21}{2} & -\frac{27}{2} & & \\
\uparrow & & & & & & &
\end{array}$$

$$\begin{array}{ccccc|cl}
1 & 2 & 1 & 0 & -1 & 5 & \\
0 & 1 & 0 & 1 & -1 & 3 & (6) \\
\hline
0 & 1 & 1 & 0 & 10 & -11 &
\end{array}$$

Here $\mathbf{c}' \geq \mathbf{0}$, so we have the final or optimum tableau for the original problem. The optimum solution is $\mathbf{x}_{opt} = (5,0,0,3,0,)^T$ and the optimum value is $f_{opt} = 11$.

4.5 Rank Deficiency, or Redundancy and Inconsistency.

If $r(\mathbf{A}) < m$ then $\mathbf{A} \not\supset \mathbf{I}_m$ so we must necessarily use the two-part simplex method and we assume, for convenience, that the artificial problem is problem II in section 4.3.

(i) If $\tilde{f}_{opt} = 0$, then the original problem is feasible, but as $r(\mathbf{A}) \neq m$ we must have an optimum tableau with the form illustrated in section 4.3 (iii) which, in this instance, we write as follows, where $\mathbf{b}_1' \geq \mathbf{0}$ and $\mathbf{b}_2' = \mathbf{0}$:

$$\left[\begin{array}{c|c|c|c|c} \begin{matrix}1 & & \\ & \ddots & \\ & & 1\end{matrix} & \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{O} & \mathbf{b}_1' \\ \hline \mathbf{O} & \mathbf{B}_3 & \mathbf{B}_4 & \begin{matrix}1 & & \\ & \ddots & \\ & & 1\end{matrix} & \mathbf{b}_2' \end{array}\right] \quad (1)$$

In practice we will not know that $r(\mathbf{A}) < m$. So we proceed as in section 4.3 (iii), pivoting on non-zero elements of $\mathbf{B}_3$. If $r(\mathbf{A}) < m$, at some stage the matrix $\mathbf{B}_3$ will be zero, in which case the last $(m-k)$ rows can be discarded and we have the situation discussed in 4.3(ii), with $m = k$. We will thus have removed the $(m - k)$ redundant equations and obtained an initial simplex tableau. We will have performed not much more work than a conventional reduction of $\mathbf{A}$ to row-echelon form to determine the redundant equations, which would only take us to the beginning of the first part.

(ii) If $\tilde{f}_{opt} > 0$, then there are no feasible solutions to the original problem, so there is no *l.p.p.* to solve. However, for a *real* problem this would probably be rather disturbing and we would wish to determine whether the given equality constraints were consistent (there are solutions but none non-negative) or inconsistent (there are no solutions at all), because this would help to determine the error in the model formulation.

The optimum tableau for problem I will be as (1) above, but now $\mathbf{b}_1' \geq \mathbf{0}$ and $\mathbf{b}_2' \geq \mathbf{0}$ and $\mathbf{b}_2' \neq \mathbf{0}$ because $\tilde{f}_{opt} > 0$ ($\mathbf{z}_2' \geq \mathbf{0}$ and $\mathbf{z}_2' \neq \mathbf{0}$). Again we proceed as in section 4.3 (iii), pivoting on any non-zero element of $\mathbf{B}_3$ and reducing the size of $\mathbf{B}_3$ until $\mathbf{B}_3$ is zero. If at that stage $\mathbf{b}_2' = \mathbf{0}$ then the original constraints

are consistent, but some are redundant. If $\mathbf{b}'_2 \neq \mathbf{0}$ then the original constraints are inconsistent; there are *no* vectors $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{b}$ (ER).

4.6

A comment is appropriate at this stage before discussing cycling. The analysis in section 4.5, along with everything that has preceded it, tacitly assumes that the arithmetic in operations ①, ②, ③ of the simplex method will be performed exactly. This will be the case in academic exercises and in certain spècial problems, but in general, when *l.p.p.s* are solved on a computer, the arithmetic is not performed exactly and so, for example, $\mathbf{B}_3$ and $\mathbf{b}'_2$ in (1) of section 4.5 will almost never be exactly zero unless they are empty throughout. A proper investigation of the consequences of this observation is a serious undertaking in the mathematical field of numerical analysis, but it leads to technical modifications in the details of the arithmetic operations rather than in the simplex method itself. (These modifications are described and discussed briefly in Appendix 3). This observation about arithmetic operations does mean, however, that section 4.5 is somewhat unrealistic for general *l.p.p.s,* unless all coefficients with magnitude less than some small number ϵ are regarded as zero. If the coefficients of $\mathbf{A}$ and $\mathbf{b}$ were given to four decimal places for example, then we might choose $\epsilon = 0.5 \times 10^{-4}$.

4.7 Cycling

This section can also be regarded as unrealistic. In practice only specially constructed *l.p.p.s* will cycle and the modification which we derive below, to the method of section 3.2(ii) for choosing s, is not incorporated in computer programs for solving *l.p.p.s.* Nevertheless it is a worthwhile excursion because the modification has a certain charm and it removes the qualifications about degenerate *b.f.s.s* in theorem 5 (section 3.5).

Cycling cannot occur if all *b.f.s.s* which the simplex method produces are non-degenerate, and if the current *b.f.s.* is non-degenerate the next one produced is degenerate if and only if there is a tie in the θ-column of the tableau. So we perturb the original problem in such a way that a tie cannot occur. For convenience we assume that $\mathbf{A} \supset \mathbf{I}_m$ so that we do not need the artificial first part.

Now replace $\mathbf{b}$ by $\mathbf{b}(\epsilon)$, where

$$\mathbf{b}(\epsilon) = \mathbf{b} + \Sigma_{j=1}^{n} \epsilon^j \mathbf{a}_{*j},$$

and ϵ is some small positive number.

In other words

$$b_i(\epsilon) = b_i + \epsilon a_{i1} + \epsilon^2 a_{i2} + \epsilon^3 a_{i3} + \ldots + \epsilon^n a_{in}, \quad i = 1, 2, \ldots, m.$$

For ϵ sufficiently small a tie cannot occur in the θ-column, whose entries are now

$$\frac{b_i(\epsilon)}{a_{it}}, \text{ for } a_{it} > 0, \; i = 1,2,\ldots,m.$$

For, suppose $\dfrac{b_1}{a_{1t}} = \dfrac{b_2}{a_{2t}} = \min\limits_{\substack{i=1,2,\ldots,m \\ a_{it}>0}} \dfrac{b_i}{a_{it}}$. Then if $\dfrac{a_{11}}{a_{1t}} > \dfrac{a_{21}}{a_{2t}}$, then $\dfrac{b_1(\epsilon)}{a_{1t}} > \dfrac{b_2(\epsilon)}{a_{2t}}$ for ϵ sufficiently small, and conversely if $\dfrac{a_{11}}{a_{1t}} < \dfrac{a_{21}}{a_{2t}}$.

If $\dfrac{a_{11}}{a_{1t}} = \dfrac{a_{21}}{a_{2t}}$, then we consider the ϵ^2 terms, $\dfrac{a_{12}}{a_{1t}}$ and $\dfrac{a_{22}}{a_{2t}}$. If these do not resolve the tie we consider the ϵ^3 terms and so on. For some j, $i \leq j \leq n$, we must have $\dfrac{a_{1j}}{a_{1t}} \neq \dfrac{a_{2j}}{a_{2t}}$ (ER). The argument clearly applies to any tableau, not just the first, and to a tie between any number of rows.

We may now observe that nowhere do we actually need the value of ϵ, so that even if we wanted to incorporate this perturbation modification to the simplex method, which we do not, we would not actually have to perturb the problem at all.

To demonstrate, we apply the modification to the example discussed in section 3.4.

The second tableau was

						$\mathbf{b}$	θ	$\mathbf{b}(\epsilon)$
2	1	−1	1	0	0	2		$2 + 2\epsilon + \epsilon^2 - \epsilon^3 + \epsilon^4$
4	0	4	1	1	0	8	2	$8 + 4\epsilon + 4\epsilon^3 + \epsilon^4 + \epsilon^5$
2	0	2	−1	0	1	4	2	$4 + 2\epsilon + 2\epsilon^3 - \epsilon^4 + \epsilon^6$
3	0	−3	2	0	0	$f+4$		
		↑						

where we have added the explicit description of $\mathbf{b}(\epsilon)$, although this is unnecessary as the coefficients involved appear in the main part of the tableau.

In this case $t = 3$ and

$$\frac{b_2}{a_{23}} = \frac{b_3}{a_{33}} = \min_{\substack{i=1,2,3\\ a_{i3}>0}} \frac{b_i}{a_{it}} \quad .$$

The ϵ terms give $\frac{4}{4} = \frac{2}{2}$, which is still a tie.
The ϵ^2 terms give $0 = 0$, which is still a tie.
The ϵ^3 terms give $\frac{4}{4} = \frac{2}{2}$, which is still a tie.
The ϵ^4 terms give $\frac{1}{4} > -\frac{1}{2}$, so $\dfrac{b_2(\epsilon)}{a_{23}} > \dfrac{b_3(\epsilon)}{a_{33}}$, so $\displaystyle\min_{\substack{i=1,2,3\\ a_{i3}>0}} \frac{b_i(\epsilon)}{a_{it}} = \frac{b_3(\epsilon)}{a_{33}}$

for ϵ sufficiently small.
Thus in this case we choose $s = 3$.

It should be emphasised that degenerate *b.f.s.s* do not necessarily lead to cycling.

An alternative approach for resolving cycling may be found in {10}.

Exercises 4

1. Solve the *l.p.p.* in canonical form defined by the following data:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 3 & -2 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \ \mathbf{c}^T = (1,1,1,1,1).$$

   Comment on the final tableau and optimum solution and obtain a different final tableau.

2. Suppose that the columns of **A** in a *l.p.p.* in canonical form include some, but not all, of the columns of $\mathbf{I}_m$. What is the artificial problem which is solved in the first part of the two-part simplex method?

   (Hint: for convenience, assume that the last k columns of **A** are the first k columns of $\mathbf{I}_m$.)

3. Solve the *l.p.p.s* below by the two-part simplex method. For the example in which $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ is unbounded (in the appropriate sense) find a feasible solution with value 1000.

   (i) *minimise* $x_1 - 2x_2 + 3x_3$ *subject to*

$$\begin{pmatrix} -2 & 1 & 3 \\ 2 & 3 & 4 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ \mathbf{x} \geq \mathbf{0}.$$

   (ii) *maximise* $3x_1 - x_2 - 3x_3 + x_4$ *subject to*

$$\begin{pmatrix} 1 & 2 & -1 & 1 \\ 2 & -2 & 3 & 3 \\ 1 & -1 & 2 & -1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 9 \\ 6 \end{pmatrix}, \ \mathbf{x} \geq \mathbf{0}.$$

   (iii) *maximise* $-x_1 + 2x_2 - x_3$ *subject to*

$$\begin{pmatrix} 3 & -2 & 0 \\ 1 & -1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}, \ \mathbf{x} \geq \mathbf{0}.$$

4. Using artificial variables find a non-negative solution of the system of equations

$$\begin{aligned} x_1 \qquad\quad - x_3 + x_4 &= 3 \\ 6x_1 - 3x_2 - x_3 \qquad &= 7 \\ 3x_1 - 2x_2 \qquad - x_4 &= 1. \end{aligned}$$

5. For the following systems of constraints, use artificial variables to find an initial simplex tableau (without the c-row) or to show that an initial simplex tableau cannot be found:

   (i) $$\begin{pmatrix} 1 & 1 & 0 & -1 & 0 \\ -1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & -1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 3 \\ 7 \end{pmatrix}, \ \mathbf{x} \geq \mathbf{0}.$$

   (ii) $$\begin{pmatrix} 1 & 1 & 0 & -1 & 0 \\ 0 & 2 & 1 & 0 & -1 \\ -2 & 4 & 3 & 2 & -3 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix}, \ \mathbf{x} \geq \mathbf{0}.$$

6. When the *l.p.p.*

$$\textit{minimise } \mathbf{c}^T \mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \text{ (where } \mathbf{b} \geq \mathbf{0})$$

is converted to canonical form, the columns of the matrix of coefficients $(\mathbf{A}, -\mathbf{I}_m)$ include no columns of $\mathbf{I}_m$ in general. Devise a simple set of $(m-1)$ row operations as a result of which only one artificial variable (instead of m) needs to be introduced in the first stage of the two-part simplex method.

7. In the ϵ perturbation method for avoiding cycling, if the initial vector $\mathbf{b}$ has some zero elements, it is possible that the *b.f.s.* defined by $\mathbf{b}(\epsilon)$ is not strictly feasible; for example, suppose $b_1 = 0$ and $a_{11} < 0$. This conflicts with the requirements for the simplex method, but is easily avoided. Explain how.

NOTES

NOTES

CHAPTER 5

DUALITY: THE DUALITY THEOREM AND CONSEQUENCES

5.1

We introduce this important, and quite unexpected, development by recalling the dietician whom we met in section 1.2. We can now imagine him happily working out an optimum diet (using the simplex method). But he is interrupted by the arrival of another character in this gastronomic drama, the nutrient tablets salesman. The salesman says that he can supply, at a price, all the nutrients in tablet form and he suggests that the dietician take advantage of this. The dietician will not be interested in this offer if the cost of so buying the component nutrients in one unit of the j-th food is greater than c_j. So the salesman's prices y_i, $i = 1, 2, \ldots, m$, of units of i-th nutrient must satisfy

$$y_1 a_{1j} + y_2 a_{2j} + \ldots + y_m a_{mj} \leq c_j, \quad j = 1, 2, \ldots, n. \tag{1}$$

In other words, the costs of artificial foods as provided by the salesman must not exceed the costs of the natural foods available to the dietician.

However, the salesman wishes his total income from any deal with the dietician to be as large as possible, so that subject to (1)

$$\text{and} \quad y_i \geq 0, \quad i = 1, 2, \ldots, m, \tag{2}$$

$$\text{he wishes to maximize} \quad y_1 b_1 + y_2 b_2 + \ldots + y_m b_m. \tag{3}$$

Thus the salesman's problem is another *l.p.p.*, namely

$$\textit{maximise} \quad \mathbf{y}^T\mathbf{b} \quad \textit{subject to} \quad \mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T, \ \mathbf{y} \geq \mathbf{0}. \tag{4}$$

This *l.p.p.* (4) is said to be the *dual problem* of the *primal problem* of section 1.2. It would seem natural to rewrite (4) transposed,

$$\text{i.e.} \quad \textit{maximise} \quad \mathbf{b}^T\mathbf{y} \quad \textit{subject to} \quad \mathbf{A}^T\mathbf{y} \leq \mathbf{c}, \ \mathbf{y} \geq \mathbf{0}, \tag{5}$$

but the form given in (4) is generally preferable.

At present the connection between the *l.p.p.s* (1) of section 1.2 and (4) is the purely formal one that they are defined in terms of the same data, but we will establish in this chapter profound general connections between the dietician's and the salesman's problems.

In fact for any *l.p.p.* there is a corresponding dual problem and

we now formally define them for *l.p.p.s* in canonical and standard forms.

5.2

(i) Canonical Form

Primal: *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},$ (1)

Dual: *maximise* $\mathbf{y}^T\mathbf{b}$ *subject to* $\mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T$ (2)

(ii) Standard Form

Primal: *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq \mathbf{0},$ (3)

Dual: *maximise* $\mathbf{y}^T\mathbf{b}$ *subject to* $\mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T,\ \mathbf{y} \geq \mathbf{0}.$ (4)

Notice that in both cases one *l.p.p.* is a minimisation problem and the other a maximisation problem, but whereas in (ii) both sets of constraints are inequalities in non-negative variables, in (i) the primal constraints are equations in non-negative variables and the dual constraints are inequalities in free variables. For this reason (ii) is called the *symmetric form* of the primal and dual *l.p.p.s* and (i) the *unsymmetric form*.

Since any *l.p.p.* can be converted to, say, canonical (primal) form two questions arise immediately: are (i) and (ii) equivalent, and what is the dual of the dual? These questions are answered by the following theorems.

Theorem 6

The canonical (unsymmetric) and standard (symmetric) forms of primal and dual *l.p.p.s* are equivalent∎

We convert the standard primal to canonical primal form.

Introducing surplus variables $z_1, \ldots, z_m$ gives

$$\textit{minimise} \quad (\mathbf{c}^T, \mathbf{0}^T)\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} \quad \textit{subject to} \quad (\mathbf{A}, -\mathbf{I}_m)\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} = \mathbf{b}, \quad \begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} \geq \mathbf{0},$$

$$\text{i.e.} \quad \textit{minimise} \quad \tilde{\mathbf{c}}^T\tilde{\mathbf{x}} \quad \textit{subject to} \quad \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}},\ \tilde{\mathbf{x}} \geq \mathbf{0}, \tag{5}$$

where $\tilde{\mathbf{c}}^T = (\mathbf{c}^T, \mathbf{0}^T)$ etc.

The *l.p.p.* (5) is precisely in canonical primal form, so its dual, by (i), is

$$\textit{maximise} \quad \tilde{\mathbf{y}}^T\tilde{\mathbf{b}} \quad \textit{subject to} \quad \tilde{\mathbf{y}}^T\tilde{\mathbf{A}} \leq \tilde{\mathbf{c}}^T. \tag{6}$$

Substituting for $\tilde{\mathbf{b}}$, $\tilde{\mathbf{c}}$ etc., we obtain

$$\textit{maximise} \quad \tilde{\mathbf{y}}^T\mathbf{b} \quad \textit{subject to} \quad \tilde{\mathbf{y}}^T(\mathbf{A}, -\mathbf{I}_m) \leq (\mathbf{c}^T, \mathbf{0}^T). \tag{7}$$

Now we put $\tilde{\mathbf{y}} = \mathbf{y}$ and (7) becomes

$$\textit{maximise} \quad \mathbf{y}^T\mathbf{b} \quad \textit{subject to} \quad \mathbf{y}^T\mathbf{A} \le \mathbf{c}^T, \ -\mathbf{y}^T \le \mathbf{0}^T,$$

i.e.
$$\textit{maximise} \quad \mathbf{y}^T\mathbf{b} \quad \textit{subject to} \quad \mathbf{y}^T\mathbf{A} \le \mathbf{c}^T, \ \mathbf{y} \ge \mathbf{0}. \tag{8}$$

The *l.p.p* (8) is exactly the standard dual defined in (ii) and hence the two definitions (i) and (ii) are really the same■

Theorem 7

The dual of the dual is the primal■

By theorem 6 we can consider either form.

We convert the canonical dual to canonical primal form as follows:

$$\textit{maximise} \quad \mathbf{y}^T\mathbf{b} \quad \textit{subject to} \quad \mathbf{y}^T\mathbf{A} \le \mathbf{c}^T, \tag{9}$$

i.e. $\textit{maximise} \quad (\mathbf{u}-\mathbf{v})^T\mathbf{b} \quad \textit{subject to} \quad (\mathbf{u}-\mathbf{v})^T\mathbf{A} \le \mathbf{c}^T, \ \mathbf{u}, \mathbf{v} \ge \mathbf{0},$

i.e. $\textit{maximise} \quad (\mathbf{u}^T, \mathbf{v}^T)\begin{pmatrix} \mathbf{b} \\ -\mathbf{b} \end{pmatrix} \quad \textit{subject to} \quad (\mathbf{u}^T, \mathbf{v}^T)\begin{pmatrix} \mathbf{A} \\ -\mathbf{A} \end{pmatrix} \le \mathbf{c}^T,$

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \ge \mathbf{0},$$

i.e. $\textit{maximise} \quad (\mathbf{u}^T, \mathbf{v}^T, \mathbf{w}^T)\begin{pmatrix} \mathbf{b} \\ -\mathbf{b} \\ \mathbf{0}_n \end{pmatrix} \quad \textit{subject to}$

$$(\mathbf{u}^T, \mathbf{v}^T, \mathbf{w}^T)\begin{pmatrix} \mathbf{A} \\ -\mathbf{A} \\ I_n \end{pmatrix} = \mathbf{c}^T, \ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} \ge \mathbf{0},$$

i.e. $\textit{minimise} \quad (\mathbf{u}^T, \mathbf{v}^T, \mathbf{w}^T)\begin{pmatrix} -\mathbf{b} \\ \mathbf{b} \\ \mathbf{0}_n \end{pmatrix} \quad \textit{subject to}$

$$(\mathbf{u}^T, \mathbf{v}^T, \mathbf{w}^T)\begin{pmatrix} \mathbf{A} \\ -\mathbf{A} \\ \mathbf{I}_n \end{pmatrix} = \mathbf{c}^T, \ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} \ge \mathbf{0},$$

i.e. $\textit{minimise} \quad \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix}^T \begin{pmatrix} -\mathbf{b} \\ \mathbf{b} \\ \mathbf{0}_n \end{pmatrix} \quad \textit{subject to}$

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix}^T \begin{pmatrix} \mathbf{A} \\ -\mathbf{A} \\ \mathbf{I}_n \end{pmatrix} = \mathbf{c}^T, \ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} \ge \mathbf{0},$$

i.e. $\textit{minimise} \quad (-\mathbf{b}^T, \mathbf{b}^T, \mathbf{0}_n^T)\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} \quad \textit{subject to}$

$$(\mathbf{A}^T, -\mathbf{A}^T, \mathbf{I}_n)\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \mathbf{c}, \ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} \ge \mathbf{0}, \tag{10}$$

i.e.
$$\textit{minimise} \quad \tilde{\mathbf{c}}^T\tilde{\mathbf{x}} \quad \textit{subject to} \quad \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}, \ \tilde{\mathbf{x}} \ge \mathbf{0}, \tag{11}$$

where $\tilde{\mathbf{c}} = \begin{pmatrix} -\mathbf{b} \\ \mathbf{b} \\ \mathbf{0}_n \end{pmatrix}$, $\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix}$, $\tilde{\mathbf{A}} = (\mathbf{A}^T, -\mathbf{A}^T, \mathbf{I}_n)$, $\tilde{\mathbf{b}} = \mathbf{c}$.

The *l.p.p* (11) is precisely in canonical primal form, so its dual is defined by (i) to be

$$\textit{maximise} \quad \tilde{\mathbf{y}}^T\tilde{\mathbf{b}} \quad \textit{subject to} \quad \tilde{\mathbf{y}}^T\tilde{\mathbf{A}} \leqq \tilde{\mathbf{c}}^T, \tag{12}$$

i.e. *maximise* $\tilde{\mathbf{y}}^T\mathbf{c}$ *subject to* $\tilde{\mathbf{y}}^T(\mathbf{A}^T, -\mathbf{A}^T, \mathbf{I}_n) \leq (-\mathbf{b}^T, \mathbf{b}^T, \mathbf{0}_n^T)$,

i.e. *minimise* $(-\tilde{\mathbf{y}})^T\mathbf{c}$ *subject to*

$$(-\tilde{\mathbf{y}})^T(\mathbf{A}^T, -\mathbf{A}^T, \mathbf{I}_n) \geq (\mathbf{b}^T, -\mathbf{b}^T, \mathbf{0}_n^T).$$

Now we put $-\tilde{\mathbf{y}} = \mathbf{x}$ because $\tilde{\mathbf{y}}$ must be an n-vector and we obtain the *l.p.p.*

minimise $\mathbf{x}^T\mathbf{c}$ *subject to* $\mathbf{x}^T\mathbf{A}^T \geq \mathbf{b}^T$, $\mathbf{x}^T(-\mathbf{A}^T) \geq -\mathbf{b}^T$, $\mathbf{x}^T \geq \mathbf{0}^T$,

i.e. *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{A}\mathbf{x} \geq \mathbf{b}$, $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$,

i.e. *minimise* $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$,

which is the canonical primal■

The proofs of theorems 6 and 7 demonstrate how to obtain the dual of *any l.p.p.* First convert it to any one of the four forms in (i) and (ii) and then use the appropriate definition (i) or (ii). It is not necessary to be quite as tedious as we have been in the proofs above, but systematic conversion of variables, constraints, and objective function is helpful, and the renaming process of (11) avoids any confusion that might be caused by quantities having the "wrong name", e.g. **y** for the variables in primal form. Any *l.p.p.* and its dual problem should together be regarded as a single entity. Given any such pair of problems it is not, strictly speaking, the case that one is the primal and one the dual; each is the dual of the other.

However, in practice, as we convert any *l.p.p.* to canonical (primal) form to solve it, it is convenient, particularly for theoretical purposes, to refer to this form of the given problem simply as *the primal,* and to the dual *l.p.p.* of the problem in this form simply as *the dual.*

We will also find it convenient to call the objective function of the dual $g(\mathbf{y})$, or simply g, thus $g(\mathbf{y}) = \mathbf{y}^T\mathbf{b}$.

5.3

The situation of the dietician and the salesman suggests, correctly, a profound connection between a *l.p.p.* and its dual problem. We now solve a simple example of a *l.p.p.* and its dual because this

will suggest, again correctly, more connections and will be useful for explaining some notation which we will need to introduce.

Primal problem:

$$\begin{aligned} &\textit{minimise} \quad f(\mathbf{x}) = 6x_1 + 8x_2 + 7x_3 + x_4 + 15x_5 \\ &\textit{subject to} \quad x_1 + x_3 + 3x_5 \geq 6, \\ &\qquad\qquad x_2 + x_3 - x_4 + x_5 \geq 5, \; x_1, \ldots, x_5 \geq 0. \end{aligned} \tag{1}$$

We introduce two surplus variables, x_6 and x_7, to convert to canonical form and since we have $\mathbf{I}_2 \subset \mathbf{A}$ we have the initial simplex tableau

$$\begin{array}{rrrrr:rr|c} 1 & 0 & 1 & 0 & 3 & -1 & 0 & 6 \\ 0 & 1 & 1 & -1 & 1 & 0 & -1 & 5 \\ \hline 6 & 8 & 7 & 1 & 15 & 0 & 0 & f \\ 0 & 0 & -7 & 9 & -11 & 6 & 8 & f-76 \\ \hline \end{array} \tag{2}$$

Two orthodox simplex stages later, we have the optimum tableau

$$\begin{array}{rrrrr:rr|c} \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} & 1 & -\frac{3}{2} & 0 & \frac{1}{2} & -\frac{3}{2} & \frac{9}{2} \\ \hline 2 & 5 & 0 & 4 & 0 & 4 & 3 & f-39 \end{array} \tag{3}$$

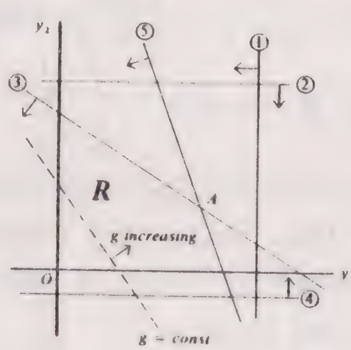
where x_3 and x_5 are the basic variables at optimality, with values $\frac{9}{2}$ and $\frac{1}{2}$ respectively.

Dual problem: this involves only two variables so we can easily solve it graphically. It is

maximise $6y_1 + 5y_2$, *subject to*

$$\begin{aligned} y_1 &\leq 6 \quad (1) \\ y_2 &\leq 8 \quad (2) \\ y_1 + y_2 &\leq 7 \quad (3) \\ -y_2 &\leq 1 \quad (4) \\ 3y_1 + y_2 &\leq 15 \quad (5) \end{aligned}$$

and $y_1, y_2 \geq 0.$



(4)

The optimum value of $g(\mathbf{y})$ occurs at the point A, which is the point $y_1 = 4$, $y_2 = 3$ and the intersection of constraint boundaries (3) and (5), and $g_{opt} = 4 \times 6 + 3 \times 5 = 39$, the same as the optimum value of the primal.

Comparing the c-row in the initial and optimum tableaux we see that the *overall row multipliers* are -4 and -3. The only way to

produce $c_1' = 2$ and $c_2' = 5$ is to add to $\mathbf{c}$ $-4\times$ (1st row of $\mathbf{A}$) and $-3\times$ (2nd row of $\mathbf{A}$), although this was done in four row-operations above. We also notice that $c_6' = 4$, $c_7' = 3$ in (3) above, corresponding to $-\mathbf{I}_2$ in columns 6 and 7 of $\mathbf{A}$.

The simplex operations to get from tableau (2) to tableau (3) are equivalent to pre-multiplication of $(\mathbf{A}, \mathbf{b})$ by $\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}$.

The columns of $\mathbf{I}_2$ in the final tableau are (in order) column 5 and column 3, and the corresponding columns of $\mathbf{A}$ provide the 2×2 matrix $\mathbf{B}$,

$$\mathbf{B} = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix} = \mathbf{B}^{-1}. \tag{5}$$

Of course $\mathbf{B}^{-1}$ must be present in the final tableau in columns 1 and 2, because these columns of $\mathbf{A}$ were $\mathbf{I}_2$.

Before producing $\mathbf{y}_{opt}$ using $\mathbf{B}^{-1}$, we might as well observe that in (3), following section 3.7,

$$\begin{aligned} \mathbf{c}'^T &= \mathbf{c}^T - 4\mathbf{e}_1{}^T\mathbf{A} - 3\mathbf{e}_2{}^T\mathbf{A} \\ &= \mathbf{c}^T - (4, 3)\mathbf{A}, \end{aligned} \tag{6}$$

$$\text{and} \quad \mathbf{B}^{-1} = \mathbf{E}_2^*\mathbf{E}_1^* = \begin{pmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{3}{2} \end{pmatrix}\begin{pmatrix} \frac{1}{3} & 0 \\ -\frac{1}{3} & 1 \end{pmatrix} (ER). \tag{7}$$

Now consider the 2-vector, $\tilde{\mathbf{c}}$ say, of original cost coefficients corresponding to columns of $\mathbf{I}_2$ (in order) in the final tableau.

We have $\tilde{\mathbf{c}} = \begin{pmatrix} c_5 \\ c_2 \end{pmatrix} = \begin{pmatrix} 15 \\ 7 \end{pmatrix}$, and

$$\tilde{\mathbf{c}}^T\mathbf{B}^{-1} = (15, 7)\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix} = (4, 3) = \mathbf{y}_{opt}^T.$$

Before we prove that not one of these observations is a coincidence, we observe that since this example is given in standard form, but solved in canonical form, it effectively illustrates both forms of the primal and dual *l.p.p.s*.

Note that the use of the notation $\tilde{\mathbf{c}}$ in the context of the solution of the dual problem is unrelated to its use elsewhere, e.g. (5) and (11) section 5.2 and problem II section 4.3.

5.4 Theorem 8. The Duality Theorem

For any *l.p.p.* and its dual problem, if either problem has an optimum solution then so does the other and the optimum values are equal ■

Corollary If either problem has *unbounded feasible solutions** then the other has no feasible solutions■

(*This is a convenient, but rather loose, abbreviation for *feasible solutions, the values of which are unbounded in the appropriate sense.*)

By theorems 6 and 7, to establish the duality theorem it is sufficient to prove that if the canonical primal problem has an optimum solution, then so has the dual and the optimum values are the same.

The duality theorem is by far the most important result in linear programming. It establishes a second dimension to the general theory and plays a prominent part in almost all applications and special methods and it, or its equivalent, appears in some quite unexpected places. Partly because of its central importance, and partly because each has its own particular interest, we shall provide three different proofs.

The first two use information from the simplex method, and are constructive proofs in that they provide an explicit expression for the solution of the dual problem. The third, by contrast, is more abstract and analytic in nature and appears in Appendix 2. It does not rely on the simplex method, but establishes the required result without providing a method for obtaining the solution of the dual. Of course, the dual problem could be solved by first converting to canonical primal form and then using the simplex method, but as we shall see, whenever the primal is solved the solution of the dual is obtainable immediately.

First Proof of the Duality Theorem

We assume that the problem is in canonical primal form and that $\mathbf{A} \supset \mathbf{I}_m$, so in practice this will usually be at the end of the first part of the two-part simplex method. Previous to this stage the tableaux have been concerned with the artificial problem, and only at the end of the first part do we obtain the initial simplex tableau for the original problem. For convenience we will denote the constraint equations at this stage by $\mathbf{Ax} = \mathbf{b}$, and we denote the first m columns of $\mathbf{A}$ by $\mathbf{B}$.

At any stage of the simplex method, let $j_1, j_2, \ldots, j_m$ be the column indices of the columns of $\mathbf{I}_m$, and denote by $\tilde{\mathbf{c}}$ the m-vector whose i-th element is the j_i-th (original) cost coefficient,

$$\text{i.e.} \quad \tilde{\mathbf{c}} = (c_{j_1}, c_{j_2}, \ldots, c_{j_m})^T.$$

Now for $j = 1, 2, \ldots, n$ denote the scalar product of this vector with the j-th column of constraint coefficients by w_j',

$$\text{i.e.}\quad w'_j = \Sigma_{i=1}^{m} (\tilde{\mathbf{c}})_i a'_{ij} = \Sigma_{i=1}^{m} c'_{j_i} a'_{ij}. \tag{1}$$

Thus $\mathbf{w}'^T$ is the n-vector $\tilde{\mathbf{c}}^T\mathbf{A}'$, the dashes indicating a general stage (see section 3.2).

So, in the tableau (2) in the previous section, for example

$$j_1 = 1, j_2 = 2, \tilde{\mathbf{c}} = (6, 8)^T,\ w_1 = 6,\ w_2 = 8,\ w_3 = 14, \ldots.$$

In the tableau (3) of the previous section, for example

$$j_1 = 5,\ j_2 = 3,\ a'_{12} = -\tfrac{1}{2},\ c'_2 = 5,$$
$$\tilde{\mathbf{c}} = (c_5, c_3)^T = (15, 7),\ w_1 = 4,\ w_2 = 3,\ w_3 = 7, \ldots$$

Lemma

At any stage of the simplex method

$$w'_j = c_j - c'_j = j\text{-th original cost coefficient } -j\text{-th } e.c.c. \ \blacksquare$$

We establish this result by induction.

Assume for convenience and *w.l.o.g.* that the $(n - m + i)$-th column of $\mathbf{A}$ is the i-th column of $\mathbf{I}_m$ for $i = 1, 2, \ldots, m$. (Thus $\mathbf{A}$ has the form $(\mathbf{B}, \ldots, \mathbf{I}_m)$.) Then the *e.c.c.s.* $\mathbf{c}'$ in the first simplex tableau (of the second stage) are defined by

$$\mathbf{c}'^T = \mathbf{c}^T - \Sigma_{i=1}^{m} c_{n-m+i} \times i\text{-th row of } \mathbf{A}$$
$$\text{i.e.}\quad c'_j = c_j - \Sigma_{i=1}^{m} c_{n-m+i}\, a_{ij},\ j = 1, 2, \ldots, n, \text{ and}$$
$$j_1 = n - m + 1, j_2 = n - m + 2, \ldots, j_m = n.$$

Also $\tilde{\mathbf{c}} = (c_{n-m+1}, \ldots, c_m)$, so that $w_j = \Sigma_{i=1}^{m} (\tilde{\mathbf{c}})_i a_{ij} = c_j - c'_j$. Thus the assertion of the lemma is true for the initial simplex tableau.

Now, using the notation established in section 3.2, and assuming *w.l.o.g.* that the first m columns of $\mathbf{A}'$ are $\mathbf{I}_m$, we show that

$$w'_j = c_j - c'_j,\ j = 1, 2, \ldots, n, \text{ implies}$$
$$w^*_j = c_j - c^*_j,\ j = 1, 2, \ldots, n.$$

It is helpful to refer back to the two tableaux pictured in section 3.2, and to recall that

$$a^*_{sj} = a'_{sj}/a'_{st},\ j = 1, 2, \ldots, n,$$
$$a^*_{ij} = a'_{ij} - a'_{it} \times a'_{sj}/a'_{st},\ j = 1, 2, \ldots, n,\ i = 1, 2, \ldots, m, i \neq s,$$
$$\text{and}\quad c^*_j = c'_j - c'_t \times a'_{sj}/a'_{st},\ j = 1, 2, \ldots, n.$$

Now, $w'_j = a'_{1j} c_1 + a'_{2j} c_2 + \ldots + a'_{sj} c_s + \ldots + a'_{mj} c_m$, because in this case $\{j_1, j_2, \ldots, j_m\} = \{1, 2, \ldots, m\}$,
and $w'_j = c_j - c'_j$ by hypothesis.

Also, $w^*_j = a^*_{1j} c_1 + a^*_{2j} c_2 + \ldots + a^*_{sj} c_t + \ldots + a^*_{mj} c_m$ because now $\{j_1, j_2, \ldots, j_s, \ldots, j_m\} = \{1, 2, \ldots, t, \ldots, m\}$.

Therefore $w_j^* = a_{1j}^* c_1 + a_{2j}^* c_2 + \ldots + 0c_s + \ldots + a_{mj}^* c_m + a_{sj}^* c_t$

$$= \left(a_{1j}^* - \frac{a'_{sj}}{a'_{st}} a'_{1t}\right) c_1 + \ldots + \left(a'_{sj} - \frac{a'_{sj}}{a'_{st}} a'_{st}\right) c_s$$

$$+ \ldots + \left(a'_{mj} - \frac{a'_{sj}}{a'_{st}} a'_{mt}\right) c_m + a_{sj}^* c_t$$

$$= \Sigma_{i=1}^m a'_{ij} c_i - \frac{a'_{sj}}{a'_{st}} \Sigma_{i=1}^m a'_{it} c_i + \frac{a'_{sj}}{a'_{st}} c_t$$

$$= w_j - \frac{a'_{sj}}{a'_{st}} (w_t - c_t) = c_j - c'_j - \frac{a'_{sj}}{a'_{st}} (-c'_t)$$

$$= c_j - \left(c'_j - \frac{a'_{sj}}{a'_{st}} c'_t\right) = c_j - c_j^* \blacksquare$$

Using this result, the proof of the duality theorem can now be completed without difficulty.

Denote the optimum solution by $\mathbf{x}_{opt}$ and assume that in the final (optimum) tableau the columns of $\mathbf{I}_m$ in order are in the first m columns of the tableau. Thus $\tilde{\mathbf{c}}^T_{opt} = (c_1, c_2, \ldots, c_m)$, and $c'_{opt} \geq \mathbf{0}$. Since the first m columns of $\mathbf{A}$ were initially the matrix $\mathbf{B}$, the simplex operations overall are equivalent to pre-multiplication by $\mathbf{B}^{-1}$, so $\mathbf{b}'_{opt} = \mathbf{B}^{-1}\mathbf{b}$, i.e. $(\mathbf{x}_{opt})_i = (\mathbf{B}^{-1}\mathbf{b})_i$, $i = 1, 2, \ldots, m$, $(\mathbf{x}_{opt})_j = 0, j = m+1, \ldots, n$.

Now consider the m-vector $\mathbf{y}_0$, where

$$\mathbf{y}_0^T = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}. \qquad (2)$$

This vector is feasible for the dual problem because

$$\mathbf{y}_0^T\mathbf{A} = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A}, \text{ and so}$$

$$(\mathbf{y}_0^T\mathbf{A})_j = \tilde{\mathbf{c}}^T\mathbf{a}'_{*j} \text{ at the optimum stage}$$

$$= w'_j \text{ at the optimum stage}$$

$$= c_j - c'_j \text{ at the optimum stage.} \qquad (3)$$

Thus $(\mathbf{y}_0^T\mathbf{A})_j - c_j = -c'_j \leq 0$,

$$\text{i.e. } \mathbf{y}_0^T\mathbf{A} \leq \mathbf{c}^T$$

The optimum value of f, $f(\mathbf{x}_{opt}) = \mathbf{c}^T\mathbf{x}_{opt} = \Sigma_{j=1}^n c_j(\mathbf{x}_{opt})_j$

$$= \Sigma_{i=1}^m c_i(\mathbf{x}_{opt})_i \text{ in this case}$$

$$= \Sigma_{i=1}^m c_i b'_i \text{ at the optimum stage}$$

$$= \Sigma_{i=1}^m c_i(\mathbf{B}^{-1}\mathbf{b})_i,$$

and the value of the objective function of the dual

$$g(\mathbf{y}_0) = \Sigma_{i=1}^m (\mathbf{y}_0^T)_i b_i = \Sigma_{i=1}^m (\tilde{\mathbf{c}}^T\mathbf{B}^{-1})_i b_i. \qquad (4)$$

Therefore $f(\mathbf{x}_{opt}) = g(\mathbf{y}_0)$.

Now, for *any* feasible solutions of the primal and dual problems $\mathbf{x}$ and $\mathbf{y}$,

$$g(\mathbf{y}) = \mathbf{y}^T\mathbf{b} = \mathbf{y}^T(\mathbf{A}\mathbf{x}) = (\mathbf{y}^T\mathbf{A})\mathbf{x} \leq \mathbf{c}^T\mathbf{x} = f(\mathbf{x}). \tag{5}$$

Hence $\mathbf{y}_0$ is in fact an optimum solution of the dual∎

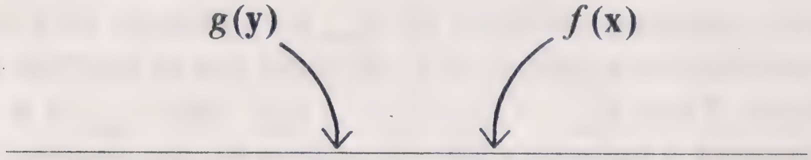
Proof of Corollary

Suppose the assertion is false, so that for any K there is a feasible solution $\mathbf{x}_K$ of the primal such that $f(\mathbf{x}_K) < K$, and suppose that the dual problem has a feasible solution $\mathbf{y}$. Then just choose $K = g(\mathbf{y})$, and as we have just seen

$$g(\mathbf{y}) \leq f(\mathbf{x}_K),$$

which is a contradiction.

It may be helpful to imagine an objective function axis on which $g(\mathbf{y})$ and $f(\mathbf{x})$ can be indicated for any feasible $\mathbf{y}$ and $\mathbf{x}$.



The primal objective function values are all to the right of any dual objective function values, from (5). These simple considerations make it clear that, apart from the vital assertion that equality must occur for optimum solutions, the duality theorem can be established from elementary considerations without the aid of the simplex method.

5.5 Consequences of the Duality Theorem

This and the following two sections are a collection of observations and further results that can be regarded as immediate consequences of theorem 8 itself and of the first proof of theorem 8. The three sections are labelled practical, theoretical and economic consequences, but these are not meant to be precise labels and it is in no sense an exhaustive collection. In fact, the duality theorem plays a prominent role in almost everything that follows. This chapter is completed with a description of the primal-dual algorithm which is a variant of the simplex method using the duality theorem directly.

Practical Consequences

It is very important to realise that the optimum solution of the dual is provided by the simplex method when solving the primal.

Referring back to (1) of the previous section and the lemma, we see that the elements of $\mathbf{y}_{opt}$ are elements of $\mathbf{w}'_{opt}$ and since $w'_j = c_j - c'_j$ we obtain elements of the optimum dual solution by subtracting appropriate *e.c.c.s.* of the final simplex stage from original cost coefficients. Now $(\mathbf{y}_{opt})_i$ is given by

$$\tilde{c}^T_{opt} \times (i\text{-th column of } \mathbf{B}^{-1}),$$

and as the initial simplex tableau (at the beginning of the second part) contains the columns of $\mathbf{I}_m$, the columns of $\mathbf{B}^{-1}$ are in the corresponding columns of the final simplex tableau. So, suppose initially the i-th column of $\mathbf{I}_m$ is the j_i-th column of $\mathbf{A}$, $i = 1, 2, \ldots, m$, then

$$(\mathbf{y}_{opt})_i = c_{j_i} - c'_{j_i}, \tag{1}$$

where the *e.c.c.s* are from the final stage.

Example 1

Consider the *l.p.p.* solved in section 3.4. Here, for the problem as given, $\mathbf{I}_m \subset \mathbf{A}$ so

$j_1 = 4,\ j_2 = 5,\ j_3 = 6,\ c_4 = 0,\ c_5 = 0,\ c_6 = 0,$

and from the final simplex tableau $c'_4 = \frac{11}{4},\ c'_5 = \frac{3}{4},\ c'_6 = 0,$

so $\mathbf{y}_{opt} = (-\frac{11}{4}, -\frac{3}{4}, 0)^T$.

Note that

$$\mathbf{y}^T_{opt}\mathbf{A} = (-7, -2, -1, -\tfrac{11}{4}, -\tfrac{3}{4}, 0),$$
$$\text{which is} \le (-1, -2, -1, \quad 0, \quad 0, 0) = \mathbf{c}^T, \tag{2}$$

and $g(\mathbf{y}_{opt}) = \mathbf{y}^T_{opt}\mathbf{b} = -\frac{22}{4} - \frac{18}{4} = -10 = \mathbf{c}^T\mathbf{x}_{opt} = f(\mathbf{x}_{opt})$.

Example 2

Consider the *l.p.p.* solved in section 4.4. Referring to tableau (5) presented there,

$j_1 = 2,\ j_2 = 4,\ c_2 = 5,\ c_4 = 2,$

and from the final tableau $c'_2 = 1,\ c'_4 = 0,$

so $\mathbf{y}_{opt} = (4, 2)^T$.

Again we note that

$$\mathbf{y}^T_{opt}\mathbf{A} = (4, 2)\begin{pmatrix} \frac{1}{2} & 1 & \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \end{pmatrix} = (1, 4, 1, 2, -3),$$

which is $\le (1, 5, 2, 2, 7)$,

$$\text{and} \quad \mathbf{y}^T_{opt}\mathbf{b} = (4, 2)\begin{pmatrix} \frac{5}{2} \\ \frac{1}{2} \end{pmatrix} = 11 = \mathbf{c}^T\mathbf{x}_{opt}. \tag{3}$$

As these two examples demonstrate, the duality theorem provides a foolproof check or verification. For any *l.p.p.* and its dual, if $\mathbf{x}_0$ and $\mathbf{y}_0$ are believed to be optimum, then if they satisfy their respective constraints and $f(\mathbf{x}_0) = g(\mathbf{y}_0)$ they are indeed optimum. When using this means of checking a calculated solution $\mathbf{x}_0$, it is vital to remember that $\mathbf{y}_0$ must satisfy the dual constraints. For example $\mathbf{y}^T = (3, 7)$, in example 2 above, gives $\mathbf{y}^T\mathbf{b} = 11$ but $(3, 7)^T$ is not an optimum solution of the dual because it does not satisfy the dual constraints.

It is equally vital to realise, in example 2 above, that the vector $\mathbf{y}_{opt}$ obtained is not the solution of the dual of the *l.p.p.* as given at the beginning of section 4.4, but only of the *l.p.p.* whose constraint equations are described by $\mathbf{A}'\mathbf{x} = \mathbf{b}'$ in section 4.4. The row operations which change the constraints from $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ to $\mathbf{A}'\mathbf{x} = \mathbf{b}'$, $\mathbf{x} \geq \mathbf{0}$ leave the solution and optimum value of the primal problem unchanged. Therefore, by the duality theorem, the optimum value of the dual problem is unchanged, but the row operations on $\mathbf{A}$ do change the dual constraints and the solution of the dual problem (see section 6.2).

Another observation, which comes directly from (5) in the previous section, is that we have a measure of *near optimality*. For any feasible vectors $\mathbf{x}$ and $\mathbf{y}$, $f(\mathbf{x}) - g(\mathbf{y})$ is the largest improvement that can be obtained for either objective function. For example, in the *l.p.p.* in canonical form in section 3.4, $\mathbf{x} = (0.1, 3.6, 1.8, 0, 0.4, 0.2)^T$ and $\mathbf{y}^T = (-2.75, -0.75, -0.25)$ both satisfy the appropriate constraints, but neither is an optimum solution. As $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = -9.1$ and $g(\mathbf{y}) = \mathbf{y}^T\mathbf{b} = -11.5$, an optimum solution for either problem will produce an improvement in the objective function value of at most 2.4. In this case both $\mathbf{x}$ and $\mathbf{y}$ are close to the optimum solutions, but this need not necessarily be the case in general when $f(\mathbf{x})$ is close to $g(\mathbf{y})$ (*ER*).

Finally, since when solving a *l.p.p.* we automatically obtain the solution of the dual problem, in practice we should solve whichever of the two will be easier when converted to canonical form. This will generally be the one which involves the smaller number of equality constraints and so, for problems in which $m > n$, one would normally solve the dual problem directly.

5.6 Theoretical Consequences

For any situation which gives rise to a *l.p.p.* the duality theorem provides a statement about that situation. These statements are sometimes celebrated results in their own right, e.g. the minimax theorem of game theory (see chapter 13) and theorems of alternatives for matrices (see chapter 6), and some, including these two, were established before the duality theorem.

The elegant and subtle result we are concerned with in this section applies to *l.p.p.s* in general, and relates zero or non-zero variables at optimality with inequality type constraints which are active or not.

Theorem 9. The Equilibrium or Complementary Slackness Theorem

For optimum solutions of the primal and dual *l.p.p.s* in standard form, the constraints of either problem corresponding to non-zero variables for the other are satisfied as equalities■

Let $\mathbf{x}_0$ and $\mathbf{y}_0$ be optimum solutions. Thus

$$\mathbf{A}\mathbf{x}_0 \geq \mathbf{b},\ \mathbf{x}_0 \geq \mathbf{0},\ \mathbf{y}_0^T\mathbf{A} \leq \mathbf{c}^T,\ \mathbf{y}_0 \geq \mathbf{0},$$
$$\text{and}\quad \mathbf{c}^T\mathbf{x}_0 = \mathbf{y}_0^T\mathbf{b}.$$

Hence
$$\mathbf{y}_0^T\mathbf{A}\mathbf{x}_0 \geq \mathbf{y}_0^T\mathbf{b} \quad\text{and}\quad \mathbf{y}_0^T\mathbf{A}\mathbf{x}_0 \leq \mathbf{c}^T\mathbf{x}_0,$$
i.e.
$$\mathbf{y}_0^T\mathbf{b} \leq \mathbf{y}_0^T\mathbf{A}\mathbf{x}_0 \leq \mathbf{c}^T\mathbf{x}_0,$$
and so
$$\mathbf{y}_0^T\mathbf{b} = \mathbf{y}_0^T\mathbf{A}\mathbf{x}_0 = \mathbf{c}^T\mathbf{x}_0,$$
and so
$$\mathbf{y}_0^T(\mathbf{b} - \mathbf{A}\mathbf{x}_0) = 0 = (\mathbf{y}_0^T\mathbf{A} - \mathbf{c}^T)\mathbf{x}_0. \tag{1}$$

Examining the left-hand equation of (1) we observe that $\mathbf{y}_0 \geq \mathbf{0}$ and $(\mathbf{b} - \mathbf{A}\mathbf{x}_0) \leq \mathbf{0}$, so that $\mathbf{y}_0^T(\mathbf{b} - \mathbf{A}\mathbf{x}_0) = 0$ if and only if

$$(\mathbf{y}_0)_i\,(\mathbf{b} - \mathbf{A}\mathbf{x}_0)_i = 0,\ i = 1, 2, \ldots, m.$$

Thus $(\mathbf{y}_0)_i > 0$ implies that $(\mathbf{A}\mathbf{x}_0)_i = b_i$.

A similar examination of the right-hand equation of (1) establishes the result■

For a general *l.p.p.* we expect the optimum solution to be non-degenerate, and so (for $m < n$) the m constraints of the dual corresponding to the m basic variables of the primal are satisfied as equalities at optimality. This provides a straightforward method of verifying that a particular vector is optimum when the "simplex record" is not available: namely, solve the m dual constraints as a system of equations, and check that the solution obtained is non-negative (for standard form) and has the same value.

Example

Consider the primal and dual problems discussed in section 5.3. The basic variables at optimality for the primal are x_3 and x_5 and the third and fifth constraints of the dual are satisfied as equations by the dual solution $\mathbf{y}_0^T = (4, 3)$.

Alternatively, with just the two *l.p.p.s* and a *suggestion* that $\mathbf{x}_0^T = (0, 0, \frac{9}{2}, 0, \frac{1}{2})$ is the primal optimum solution, the third and fifth dual constraints are $y_1 + y_2 \leq 7$, $3y_1 + y_2 \leq 15$. Solving the equations

$$\begin{pmatrix} 1 & 1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 7 \\ 15 \end{pmatrix} \tag{2}$$

gives $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$. This value of $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is non-negative and satisfies *all* the constraints and $(4, 3)\begin{pmatrix} 6 \\ 5 \end{pmatrix} = 39$, verifying that $\mathbf{x}_0$ is optimum.

Using the problem of section 5.3 to illustrate the equilibrium theorem suggests, correctly, that there is a corresponding result for the primal and dual problems in canonical form (see exercise 5.8).

It is worth mentioning here, for those who have met Lagrange multipliers, that the dual variables are just the Lagrange multipliers for the problem

$$\textit{minimise}\, f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},$$

and that the equilibrium theorem corresponds to the Kuhn-Tucker conditions of non-linear optimisation (see {10}, {12}, {13}, {14}).

5.7 Economic Consequences

The basic point to be made is that the dual problem usually has a meaningful interpretation which provides useful, sometimes crucial, insights into the original problem. The dietician and the salesman with whom this chapter started are a good example. The dietician now knows that accepting the salesman's offer will not produce a cheaper diet than one already available, and the salesman now knows that however he chooses his prices he cannot ensure for himself more than a certain fixed return. But more precise statements can be made connecting the real or natural economy of the dietician and the alternative or synthetic economy offered by the salesman when they are running optimally.

For example, from the equilibrium theorem we see that the dietician supplies none of any food that is overpriced compared to its synthetic equivalent, i.e. of any food whose "real price" exceeds its "shadow

price". Also, the salesman will supply free of charge any nutrient whose requirement is exceeded in the actual diet.

The dual of the transportation problem is left until chapter 10, but can be referred to now (see exercise 10.1).

The dual of the manufacturer's problem, exercise 1.5, is in standard primal form and can be interpreted as the problem of a rival manufacturer with a takeover bid. The rival offers to buy the resources and wishes to do so at minimum cost. His offer will be unacceptable if the manufacturer would obtain more money by using his resources to continue to make the products. Here, for example, the rival will set a price of zero for any resource in surplus. Alternatively the dual variables provide a set of replacement prices for the resources which will exactly exhaust the manufacturer's profit and $(\mathbf{y}_{opt})_i$ can therefore be interpreted as the *implied* or *imputed* value of one unit of the i-th resource.

The relationship $\mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{b}$, where $\mathbf{x}$ is the current *b.f.s.*, is satisfied at every stage of the simplex method if $\mathbf{y}^T = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}$ and $\tilde{\mathbf{c}}$ and $\mathbf{B}^{-1}$ are also given their current values (see exercise 6.9). As the elements of $\mathbf{x}$ represent a feasible set of operating levels for the manufacturer's n activities, the elements of $\mathbf{y}$ are the implied resource values corresponding to this manufacturing programme, so their interpretation as shadow prices is valid at all stages. For the optimum manufacturing programme, the equilibrium theorem shows that no activity is operated at a positive level if the activity would lose money if the resources used were costed at their implied prices. Also, if a disposal activity (slack variable) operates at a positive level the implied value of the corresponding resource is zero.

The m-vector $\mathbf{y}$ is sometimes denoted by $\boldsymbol{\pi}$ and its elements called *simplex multipliers* or *pricing multipliers*. This is because the elements of $\mathbf{y}$ at any stage provide the multiples of the rows of $\mathbf{A}$ that have been subtracted from the cost coefficients $\mathbf{c}$ to obtain the *e.c.c.s* $\mathbf{c}'$ (see section 6.1). (Remember the observation in section 5.5, that the simplex multipliers $\tilde{\mathbf{c}}^T\mathbf{B}^{-1}$ are just appropriate elements of the n-vector $\mathbf{w}$, and also that, as the manufacturer's problem is in standard dual form, the introduction of m disposal activities takes us immediately to part II with the original matrix of coefficients $\mathbf{A}$ unchanged.)

The economic model of the manufacturer's situation contains several ideas used to develop interindustry models of (e.g. national) economies using input-output analysis. Suppose the economy is divided into n industries or sectors, and in any one accounting period x_j is the number

of units of output of the j-th industry, $j = 1, 2, \ldots, n$. If a_{ij} is the number of units of the i-th industry's output used to produce one unit of the j-th industry's output, then $\mathbf{b} = \mathbf{x} - \mathbf{A}\mathbf{x} = (\mathbf{I} - \mathbf{A})\mathbf{x}$ gives the number of units of the output of each industry available for non-productive consumption. The elements of the matrix $\mathbf{A}$ are input-output coefficients and $(\mathbf{I} - \mathbf{A})$ is known as a *Leontief matrix.* Thus a production vector $\mathbf{x}$ which results in a specified final-demand vector $\mathbf{b}$ satisfies

$$(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0},$$

or

$$(\mathbf{I} - \mathbf{A})\mathbf{x} \geq \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0},$$

and in practice would also have to satisfy certain minimum or maximum production constraints for certain industries, $\mathbf{x} \geq \mathbf{u}$ and $\mathbf{x} \leq \mathbf{v}$. The coefficients of the objective function, whether to be maximised or to be minimised in the problem being examined, are chosen to reflect the particular objective involved.

In an economic context the inverse of a square $(m \times m)$ matrix of coefficients is often of interest and we note that such an inverse is provided by the final simplex tableau (see also exercise 7.3).

5.8 The Primal-Dual Algorithm

This variant of the simplex method is more efficient for certain particular forms of *l.p.p.* and is a way of using the dual problem and the duality theorem directly to aid progress towards the optimum solution. The idea is to satisfy the conditions of the equilibrium or complementary slackness theorem (section 5.6) using slackness in the dual to make efficient choices of primal variables to become basic.

We shall establish the algorithm from a theoretical point of view and use an example to illustrate some of the practical details. The algorithm solves the *l.p.p.*

$$P: \quad \textit{minimise } \mathbf{c}^T\mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0}$$

in the case $\mathbf{A} \not\supset \mathbf{I}_m$ without the artificial first part of the two-part simplex method, although there are still up to m extra variables involved. The savings over the two-part simplex method that might be expected are not always realised in practice so the algorithm is presented here as an interesting application of the duality results rather than as an improved method. *There are no consequences for later chapters if this section is omitted.*

Suppose we have a vector $\mathbf{y}$ satisfying the constraints $\mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T$ of the dual problem

$$D\text{: } \textit{maximise } \mathbf{y}^T\mathbf{b} \textit{ subject to } \mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T$$

and we denote by J the set of constraint indices for which equality occurs.

Thus $j \in J$ if $(\mathbf{y}^T\mathbf{A})_j = c_j$,
and $j \notin J$ if $(\mathbf{y}^T\mathbf{A})_j < c_j$, $j = 1, 2, \ldots, n$.

Suppose now that $\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix}$ is an optimum solution of the *l.p.p.*

ARP: minimise $\Sigma_{i=1}^{m} v_i$ *subject to* $\mathbf{A}\mathbf{x} + \mathbf{v} = \mathbf{b}$, $\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} \geq \mathbf{0}$, $x_j = 0$ *if* $j \notin J$,

and suppose also that $\mathbf{v} = \mathbf{0}$. Then $\mathbf{x}$ and $\mathbf{y}$ are optimum solutions of P and D because $\mathbf{A}\mathbf{x} + \mathbf{v} = \mathbf{b}$ and $\mathbf{v} = \mathbf{0}$ implies that $\mathbf{A}\mathbf{x} = \mathbf{b}$, and so $\mathbf{x}$ is feasible for P and

$$\begin{aligned}\mathbf{y}^T\mathbf{b} &= \mathbf{y}^T\mathbf{A}\mathbf{x} = \Sigma_{j=1}^{n} (\mathbf{y}^T\mathbf{A})_j x_j = \Sigma_{j\in J} (\mathbf{y}^T\mathbf{A})_j x_j = \Sigma_{j\in J} c_j x_j\\ &= \Sigma_{j=1}^{n} c_j x_j = \mathbf{c}^T\mathbf{x}.\end{aligned}$$

The *l.p.p. ARP* may be written

$$\textit{minimise } (\mathbf{0}^T, \mathbf{e}^T)\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} \textit{ subject to } (\mathbf{A}, \mathbf{I}_m)\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} = \mathbf{b},\ \begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} \geq \mathbf{0},$$
$$\mathbf{K}_J\mathbf{x} = \mathbf{0},$$

where $\mathbf{e}^T = (1, 1, \ldots, 1)$ and $\mathbf{K}_J$ is the $(n-k) \times n$ matrix whose rows are the unit vectors $\mathbf{e}_j$ for $j \notin J$, and k is the number of integers in the set J. This *l.p.p.* is called the *Associated Restricted Primal* and is the artificial problem of the two-part simplex method with the extra constraints $\mathbf{K}_J\mathbf{x} = \mathbf{0}$. The vector $\mathbf{y}$ involved in the definition of *ARP* improves at each stage, i.e. as $\mathbf{y}^T\mathbf{b}$ increases, and as the number of integers in J increases so $\mathbf{K}_J$ decreases in size. The dual of *ARP*, called the *Associated Restricted Dual,* is

$$ARD\text{: } \textit{maximise } \mathbf{u}^T\mathbf{b} \textit{ subject to } (\mathbf{u}^T\mathbf{A})_j \leq 0,\ j \in J,\ \mathbf{u} \leq \mathbf{e}\ (ER).$$

One stage of the algorithm starts with a current vector $\mathbf{y}'$ say, which is feasible for D, solves *ARP*, and hence solves *ARD* and uses this dual solution to produce $\mathbf{y}^*$ say, which is feasible for D and which satisfies

$$\mathbf{y}^{*T}\mathbf{b} > \mathbf{y}'^T\mathbf{b}.$$

So we denote by $\begin{pmatrix}\mathbf{x}'\\ \mathbf{v}'\end{pmatrix}$ the optimum solution of *ARP*, where $\mathbf{v}' \neq \mathbf{0}$ or $\mathbf{x}'$ is optimum for P, and denote by $\mathbf{u}'$ the optimum solution of *ARD*.

If $\mathbf{u}'^T\mathbf{A} \leq \mathbf{0}^T$, i.e. $(\mathbf{u}'^T\mathbf{A})_j \leq 0$, $j = 1, 2, \ldots, n$, then P is infeasible, because for any $\alpha > 0$

$$(\mathbf{y}' + \alpha\mathbf{u}')^T\mathbf{A} = \mathbf{y}'^T\mathbf{A} + \alpha\mathbf{u}'^T\mathbf{A} \leq \mathbf{c}^T + \alpha\mathbf{u}'^T\mathbf{A} \leq \mathbf{c}^T,$$

and $(\mathbf{y}' + \alpha\mathbf{u}')^T\mathbf{b} = \mathbf{y}'^T\mathbf{b} + \alpha\mathbf{u}'^T\mathbf{b} = \mathbf{y}'^T\mathbf{b} + \alpha\mathbf{e}^T\mathbf{v}'$,

so that there are feasible vectors for D with values unbounded above. Assuming P is feasible, there must be some j for which $(\mathbf{u}'^T\mathbf{A})_j > 0$, and we define

$$\theta' = \min_{\substack{j=1,2,\ldots,n \\ (\mathbf{u}'^T\mathbf{A})_j>0}} \left\{ \frac{c_j - (\mathbf{y}'^T\mathbf{A})_j}{(\mathbf{u}'^T\mathbf{A})_j} \right\} = \frac{c_t - (\mathbf{y}'^T\mathbf{A})_t}{(\mathbf{u}'^T\mathbf{A})_t} \quad \text{say.}$$

For θ small enough and positive,

$$\mathbf{y}(\theta)^T\mathbf{A} = (\mathbf{y}' + \theta\mathbf{u}')^T\mathbf{A} = \mathbf{y}'^T\mathbf{A} + \theta\mathbf{u}'^T\mathbf{A} \leq \mathbf{c}^T,$$

because $(\mathbf{u}'^T\mathbf{A})_j = 0$ for $(\mathbf{y}^T\mathbf{A})_j = c_j$, and

$$\mathbf{y}(\theta)^T\mathbf{b} = \mathbf{y}'^T\mathbf{b} + \theta\mathbf{u}'^T\mathbf{A} > \mathbf{y}'^T\mathbf{b}.$$

Thus $\mathbf{y}(\theta)$ satisfies the constraints of D and $\mathbf{y}(\theta)^T\mathbf{b}$ increases as θ increases. The maximum value that we can give to θ is θ' and for this value we have $\mathbf{y}^* = \mathbf{y}' + \theta'\mathbf{u}'$. The set J is changed because the column index t has to be included, but column indices corresponding to positive x_j remain in $J(ER)$. This gives a new ARP and takes us to the beginning of the next stage.

It is important to see that solving ARP at each stage consists simply of a normal simplex stage for P with x_t becoming basic. The vector $\begin{pmatrix}\mathbf{x}' \\ \mathbf{v}'\end{pmatrix}$ is feasible for the *next* ARP because the constraints are just those of the current ARP except that one is removed, and the only way $\begin{pmatrix}\mathbf{x}' \\ \mathbf{v}'\end{pmatrix}$ could be improved is to increase x_t from its current value of zero. Provided that increasing x_t does reduce the value of the new ARP from that given by $\begin{pmatrix}\mathbf{x}' \\ \mathbf{y}'\end{pmatrix}$, then the pivotal operations ① and ② of the simplex method (section 3.2) will optimise the new ARP. The operations are equivalent to a simplex stage for P (without calculating $\mathbf{c}'$) with a different motivation for the choice of pivotal column. To prove that the value of the ARP decreases it is sufficient to prove that in the *e.c.c.s.* for the ARP, $c_t' < 0$. Now the previous stages, as they consist just of row operations on the coefficients $(\mathbf{A}, \mathbf{I}_m\mathbf{b})$, can be represented by pre-multiplication by a matrix $\mathbf{B}^{-1}$, so the *e.c.c.s.* for ARP are given by

$$\mathbf{c}'^T = \mathbf{c}^T - \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A},$$

where $\mathbf{c}^T = (\mathbf{0}^T, \mathbf{e}^T)$ and $\tilde{\mathbf{c}}^T$ is as defined in section 5.4.

As we have seen, $\tilde{\mathbf{c}}^T\mathbf{B}^{-1}$ gives the dual variables at optimality, so that $\mathbf{c} - \mathbf{c}' = \mathbf{u}'^T\mathbf{A}$, and in particular $c_t - c_t' = (\mathbf{u}'^T\mathbf{A})_t$, and hence $c_t \leq 0$.

A similar argument to that used for theorem 5 (section 3.5) can be used to show that the primal-dual algorithm reaches the optimum solution in a finite number of stages.

The example below shows that the stages of the primal-dual algorithm can be performed in a sequence of tableaux similar to those of the simplex method.

In this situation the rows labelled $-\mathbf{u}'^T\mathbf{A}$ correspond to the c'-rows of the simplex method and are obtained in the same way, as are the rows of the following tableau once the pivotal element has been chosen. The extra row labelled $\mathbf{c}^T = \mathbf{y}'^T\mathbf{A}$ is included for clarity and so is the *θ-row* of ratios $(c_j - (\mathbf{y}^T\mathbf{A})_j)/(\mathbf{u}^T\mathbf{A})_j$, which of course, must not be confused with the *θ-column* of ratios which determine the pivotal row.

Beneath each tableau are listed the various quantities mentioned in the description of the algorithm. Notice that for examples in which $\mathbf{A} \geq \mathbf{0}$ and $\mathbf{c} \geq \mathbf{0}$, which is the case here, the feasible solution of the dual needed to start the algorithm is given by $\mathbf{y} = \mathbf{0}$.

Minimise $x_1 + 2x_2 + x_3 + 6x_4$

subject to $\begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & 1 & 3 & 1 \end{pmatrix}\mathbf{x} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, $\mathbf{x} \geq \mathbf{0}$.

								θ
	1	1	1	2	1	0	3	$\frac{3}{1}$
	2	1	3	1	0	1	4	$\frac{4}{3}$ ←
	0	0	0	0	1	1		
$-\mathbf{u}'^T\mathbf{A}$	−3	−2	−4	−3	0	0	−7	
$\mathbf{c}^T - \mathbf{y}'^T\mathbf{A}$	1	2	1	6				
θ	$\frac{1}{3}$	1	$\frac{1}{4}$	2				
			↑					

$\mathbf{y}^T = (0, 0)$, $\mathbf{y}'^T\mathbf{A} = (0, 0, 0, 0)$, J is empty

$\begin{pmatrix} \mathbf{x}' \\ \mathbf{v}' \end{pmatrix} = (0, 0, 0, 0, 3, 4)$, *value* $= 7$,

$\mathbf{u}'^T = (1, 1)$, $\mathbf{u}'^T\mathbf{A} = (3, 2, 4, 3)$, $\mathbf{u}'^T\mathbf{b} = 7$,

$\theta = \frac{1}{4}$, $\mathbf{y}^* = (\frac{1}{4}, \frac{1}{4})$, $\mathbf{y}^{*T}\mathbf{b} = \frac{7}{4}$.

$$
\begin{array}{r|cccc|cc|c|l}
 & & & & & & & & \theta \\
 & \frac{1}{3} & \frac{2}{3} & 0 & \frac{5}{3} & 1 & -\frac{1}{3} & \frac{5}{3} & 5 \\
 & \frac{2}{3} & \frac{1}{3} & 1 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{4}{3} & 2 \leftarrow \\
\hline
-\mathbf{u}'^T\mathbf{A} & -\frac{1}{3} & -\frac{2}{3} & 0 & -\frac{5}{3} & 0 & \frac{4}{3} & -\frac{5}{3} & \\
\mathbf{c}^T - \mathbf{y}'^T\mathbf{A} & \frac{1}{4} & \frac{3}{2} & 0 & \frac{21}{4} & & & & \\
\hline
\theta & \frac{3}{4} & \frac{9}{4} & & \frac{63}{20} & & & & \\
 & \uparrow & & & & & & &
\end{array}
$$

$\mathbf{y}'^T = (\frac{1}{4}, \frac{1}{4})$, $\mathbf{y}'^T\mathbf{A} = (\frac{3}{4}, \frac{1}{2}, 1, \frac{3}{4})$, $J = \{3\}$,

$\begin{pmatrix}\mathbf{x}' \\ \mathbf{v}'\end{pmatrix} = (0, 0, \frac{4}{3}, 0, \frac{5}{3}, 0)$, *value* $= \frac{5}{3}$,

$\mathbf{u}'^T = (1 - 0, 1 - \frac{4}{3}) = (1, -\frac{1}{3})$, $\mathbf{u}'^T\mathbf{A} = (\frac{1}{3}, \frac{2}{3}, 0, \frac{5}{3})$,

$\theta = \frac{1}{4}$, $\mathbf{y}^* = (1, 0)$, $\mathbf{y}^{*T}\mathbf{b} = 3$.

$$
\begin{array}{r|cccc|cc|c|l}
 & & & & & & & & \theta \\
 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{3}{2} & 1 & -\frac{1}{2} & 1 & 2 \leftarrow \\
 & 1 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 2 & 4 \\
\hline
-\mathbf{u}'^T\mathbf{A} & 0 & -\frac{1}{2} & \frac{1}{2} & -\frac{3}{2} & 0 & \frac{3}{2} & -1 & \\
\mathbf{c}^T - \mathbf{y}'^T\mathbf{A} & 0 & 1 & 0 & 4 & & & & \\
\hline
\theta & & 2 & & \frac{8}{3} & & & & \\
 & & \uparrow & & & & & &
\end{array}
$$

$\mathbf{y}'^T = (1, 0)$, $\mathbf{y}'^T\mathbf{A} = (1, 1, 1, 2)$, $J = \{1, 3\}$,

$\begin{pmatrix}\mathbf{x}' \\ \mathbf{v}'\end{pmatrix} = (2, 0, 0, 0, 1, 0)$, *value* $= 1$,

$\mathbf{u}'^T = (1, -\frac{1}{2})$, $\mathbf{u}'^T\mathbf{A} = (0, \frac{1}{2}, -\frac{1}{2}, \frac{3}{2})$, $\mathbf{u}'^T\mathbf{b} = 1$,

$\theta = 2$, $\mathbf{y}^* = (3, -1)$, $\mathbf{y}^{*T}\mathbf{b} = 5$.

$$
\begin{array}{cccc|cc|c}
0 & 1 & -1 & 3 & 2 & -1 & 2 \\
1 & 0 & 2 & -1 & -1 & 1 & 1 \\
\hline
0 & 0 & 0 & 0 & 1 & 1 &
\end{array}
$$

$\mathbf{y}'^T = (3, -1)$, $\mathbf{y}'^T\mathbf{A} = (1, 2, 0, 5)$, $J = \{2, 1\}$,

$\begin{pmatrix}\mathbf{x}' \\ \mathbf{y}'\end{pmatrix} = (1, 2, 0, 0, 0, 0)$, *value* $= 5$.

In the last (incomplete) tableau, the optimum solution $\begin{pmatrix}\mathbf{x}' \\ \mathbf{y}'\end{pmatrix}$ of *ARP* has $\mathbf{v}' = \mathbf{0}$, so we have the optimum solution of the primal problem given by $\mathbf{x} = (1, 2, 0, 0)^T$. Notice that the optimum solution of *ARD*, $\mathbf{u}'$, is obtained at each stage by subtracting the elements in the $-\mathbf{u}'^T\mathbf{A}$ row corresponding to $\mathbf{v}$ from the vector $\mathbf{e}$. This is just

applying the formula of (1) section 5.5, or (6) section 6.1, to *ARP* and *ARD*, where $\mathbf{c}^T = (\mathbf{0}^T, \mathbf{e}^T)$.

Exercises 5

1. Establish theorem 6 by starting with the canonical primal and converting it to standard primal form.
2. Establish theorem 7 using standard form instead of canonical form.
3. Obtain the dual of the *l.p.p.*

$$\textit{maximise } \mathbf{c}^T\mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b}.$$

4. For the *l.p.p.*

$$\textit{minimise } x_2 - 3x_3 + 2x_5 \textit{ subject to } x_1, x_2, \ldots, x_6, \geq 0$$

$$\textit{and} \quad \begin{pmatrix} 1 & 3 & -1 & 0 & 2 & 0 \\ 0 & -2 & 4 & 1 & 0 & 0 \\ 0 & 4 & 3 & 0 & 8 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 7 \\ 12 \\ 10 \end{pmatrix}$$

verify all the features of proof 1 of the duality theorem:

$$\tilde{\mathbf{c}}, j_1, j_2, j_3, \mathbf{w}, \mathbf{B}, \mathbf{B}^{-1}, \mathbf{x}_{opt}, \mathbf{y}_{opt}, \text{etc.}$$

Simplex tableaux for this problem are as follows:

$$\begin{array}{cccccc|c|c}
 & & & & & & & \theta \\
1 & 3 & -1 & 0 & 2 & 0 & 7 & \\
0 & -2 & 4 & 1 & 0 & 0 & 12 & \frac{12}{4} \leftarrow \\
0 & -4 & 3 & 0 & 8 & 1 & 10 & \frac{10}{3} \\
\hline
0 & 1 & -3 & 0 & 2 & 0 & f & \\
 & & \uparrow & & & & &
\end{array}$$

$$\begin{array}{cccccc|c|c}
1 & \frac{5}{2} & 0 & \frac{1}{4} & 2 & 0 & 10 & \frac{20}{5} \leftarrow \\
0 & -\frac{1}{2} & 1 & \frac{1}{4} & 0 & 0 & 3 & \\
0 & -\frac{5}{2} & 0 & -\frac{3}{4} & 8 & 1 & 1 & \frac{2}{5} \\
\hline
0 & -\frac{1}{2} & 0 & \frac{3}{4} & 2 & 0 & 9 & \\
 & \uparrow & & & & & &
\end{array}$$

$$\begin{array}{cccccc|c|}
\frac{2}{5} & 1 & 0 & \frac{1}{10} & \frac{4}{5} & 0 & 4 \\
\frac{1}{5} & 0 & 1 & \frac{3}{10} & \frac{2}{5} & 0 & 5 \\
1 & 0 & 0 & -\frac{1}{2} & 10 & 1 & 11 \\
\hline
\frac{1}{5} & 0 & 0 & \frac{4}{5} & \frac{12}{5} & 0 & 11
\end{array}$$

5. For the *l.p.p.s* of exercises 3.1(i) and 3.1(ii), write down the solution of the dual problem and use the duality theorem to verify that the primal solution *is* optimum.
6. For the *l.p.p.* in 4 above, write down the matrix operations of each stage, and hence write the solution of the dual in elementary product form (see section 3.7 and equation (2) of section 5.4).

7. Prove that

   primal infeasible does not imply dual has unbounded feasible solutions.

   Hint: construct a counter example with, for example, $m = n = 2$ and primal constraints inconsistent.
8. State and prove, for the primal and dual *l.p.p.s* in canonical form, a theorem corresponding to the equilibrium theorem.
9. For the *l.p.p.* (in standard dual form)

   $$\textit{maximise} \quad 2x_1 + 4x_2 + x_3 + x_4 \quad \textit{subject to} \quad x_1, x_2, x_3, x_4 \geq 0$$

   $$\textit{and} \quad \begin{pmatrix} 1 & 3 & 0 & 1 \\ 2 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 \end{pmatrix} \mathbf{x} \leq \begin{pmatrix} 4 \\ 3 \\ 3 \end{pmatrix}$$

   use the equilibrium theorem to verify that $\mathbf{x} = (1, 1, \frac{1}{2}, 0)^T$ is the optimum solution.
10. In the case of a degenerate optimum solution of a *l.p.p.* in standard primal form, are dual constraints corresponding to basic variables with value zero active or not?
11. In the light of the duality theorem discuss the salesman's objective and whether he should modify it in order to sell his products. Discuss also the "package-deal" aspect of the dietician/salesman and the manufacturer/rival situations.
12. Solve the *l.p.p.* below by the primal-dual algorithm, and note that the two-part simplex method would involve more work.

    $$\textit{Minimise} \quad 2x_1 + 4x_2 + 3x_3 + 3x_4 \quad \textit{subject to}$$

    $$\begin{pmatrix} 4 & 3 & 1 & 4 \\ 2 & 2 & 5 & 3 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 11 \\ 6 \end{pmatrix}, \quad \mathbf{x} \geq \mathbf{0}.$$
13. The description in section 5.8 of the primal-dual algorithm shows that at each stage a column index t is chosen to enter the set J. Does this imply that the algorithm always finds the optimum solution in at most n stages (assuming nondegeneracy)?

NOTES

CHAPTER 6

DUALITY CONTINUED: A MATRIX VIEW OF THE DUALITY THEOREM; THEOREMS OF ALTERNATIVES

6.1 Second Proof of the Duality Theorem

This is really a compact version of the first proof, and takes an overview of the whole simplex process.

We again assume that the *l.p.p.* is in canonical form with $\mathbf{A} \supset \mathbf{I}_m$, that the problem is solved by the simplex method and that at optimality $x_1, x_2, \ldots, x_m$ are the basic variables corresponding to $\mathbf{I}_m$ in the first m columns of the final tableau.

As we have done before, we partition $\mathbf{A}$, $\mathbf{x}$ and $\mathbf{c}$ into an $m \times m$ and an $m \times (n-m)$ matrix and m- and $(n-m)$-vectors respectively, and we denote the first m columns of $\mathbf{A}$ by $\mathbf{B}$. Thus the *l.p.p.* is

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0},$$

which becomes

$$\textit{minimise} \quad \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2 \quad \textit{subject to}$$

$$\mathbf{B}\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = (\mathbf{B}, \mathbf{A}_2)\mathbf{x} = \mathbf{b}, \ \begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix} \geq \mathbf{0}. \tag{1}$$

Since the overall effect of the simplex stages is to reduce $\mathbf{B}$ to $\mathbf{I}_m$, the constraint equations as given at optimality must be

$$(\mathbf{I}_m, \mathbf{B}^{-1}\mathbf{A}_2)\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix} = \mathbf{b}' = \mathbf{B}^{-1}\mathbf{b},$$

and the optimum solution is given by

$$\mathbf{x}_1 = \mathbf{b}' = \mathbf{B}^{-1}\mathbf{b}, \ \mathbf{x}_2 = \mathbf{0}. \tag{2}$$

The crucial step is to identify the vector $\mathbf{c}'_{opt}$ of *e.c.c.s* at optimality. This is obtained by subtracting from $\mathbf{c}$ multiples of rows of $\mathbf{A}$, which is equivalent to subtracting from $\mathbf{c}$ multiples of $\mathbf{B}^{-1}\mathbf{A}$, the coefficient matrix in the optimal tableau. The defining property of $\mathbf{c}'_{opt}$ is that $(\mathbf{c}'_{opt})_i = 0$, $i = 1, 2, \ldots, m$, so that

$$\mathbf{c}'^T_{opt} = \mathbf{c}^T - \Sigma_{i=1}^m c_i \times (i\text{-th row of } \mathbf{B}^{-1}\mathbf{A}). \tag{3}$$

Thus

$$\mathbf{c}'^T_{opt} = \mathbf{c}^T - \mathbf{c}_1^T(\mathbf{B}^{-1}\mathbf{A}),\ \mathbf{c}'^T_{opt} = \mathbf{c}^T - (\mathbf{c}_1^T\mathbf{B}^{-1})\mathbf{A} \geq \mathbf{0}^T, \tag{4}$$

as $\mathbf{c}' \geq \mathbf{0}$ is the optimality criterion of the simplex method.

Alternatively, we can remember that the *e.c.c.s* express the objective function in terms of non-basic variables, so eliminating $\mathbf{x}_1$ from $f = \mathbf{c}^T\mathbf{x}$ gives

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{c}^T\mathbf{x} = \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2 \\ &= \mathbf{c}_1^T(\mathbf{B}^{-1}\mathbf{b} - \mathbf{B}^{-1}\mathbf{A}_2\mathbf{x}_2) + \mathbf{c}_2^T\mathbf{x}_2 \qquad (5) \\ &= \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{b} + \mathbf{0}^T\mathbf{x}_1 + (\mathbf{c}_2^T - \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{A}_2)\mathbf{x}_2, \\ \text{so that}\quad \mathbf{c}'^T_{opt} &= (\mathbf{0}^T, \mathbf{c}_2^T - \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{A}_2) \\ &= (\mathbf{c}_1^T - \mathbf{c}_1^T,\ \mathbf{c}_2^T - \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{A}_2) \\ &= (\mathbf{c}_1^T - \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{B},\ \mathbf{c}_2^T - \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{A}_2) \\ &= \mathbf{c}^T - \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{A} \quad \text{again, and at optimality} \\ f(\mathbf{x}) &= \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{b}, \ \text{because}\quad \mathbf{x}_2 = \mathbf{0}. \end{aligned}$$

As in the first proof we now consider the vector $\mathbf{y}_0^T = \mathbf{c}_1^T\mathbf{B}^{-1}$ (notice that the vector $\tilde{\mathbf{c}}^T$ in the first proof, which is defined there at every stage of the simplex method, is here just

$$\mathbf{c}_1^T = (c_1, c_2, \ldots, c_m) \quad \text{at the optimum stage).}$$

Now $\mathbf{y}_0$ is feasible for the dual problem because

$$\mathbf{y}_0^T\mathbf{A} = \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{A} \leq \mathbf{c}^T,$$

and $g(\mathbf{y}_0) = \mathbf{y}_0^T\mathbf{b} = \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{b} = f(\mathbf{x}_{opt})$, and thus, using the argument from (5) of section 5.4, we establish the duality theorem.

It is worthwhile recalling the observation near the beginning of section 5.5. The solution of the dual is in fact given by

$$\mathbf{y}_0^T = \tilde{\mathbf{c}}^T - \tilde{\mathbf{c}}'^T_{opt}, \quad \text{where} \quad (\tilde{\mathbf{c}})_i = c_{j_i},$$

and j_i is the column index in $\mathbf{A}$ of the i-th column of $\mathbf{I}_m$.

We also recall the n-vector $\mathbf{w}'$ of section 5.4 and observe that at any stage of the simplex method, using $\mathbf{B}$ to denote the $m \times m$ matrix of columns of $\mathbf{A}$ corresponding to basic variables,

$$\mathbf{w}'^T = \tilde{\mathbf{c}}^T\mathbf{A}' = \tilde{\mathbf{c}}^T(\mathbf{B}^{-1}\mathbf{A}) = (\tilde{\mathbf{c}}^T\mathbf{B}^{-1})\mathbf{A} = \mathbf{c}^T - \mathbf{c}'^T$$

so that $\mathbf{c}'^T = \mathbf{c}^T - (\tilde{\mathbf{c}}^T\mathbf{B}^{-1})\mathbf{A}$.

Thus, with a suitable interpretation, the equations of (3) and (4) hold throughout the simplex method and the vector $\tilde{\mathbf{c}}^T\mathbf{B}^{-1}$ gives the multiples of the rows of $\mathbf{A}$ that have been subtracted from $\mathbf{c}^T$ to obtain the *e.c.c.s* $\mathbf{c}'^T$.

A comparison of this section and section 5.4 demonstrates very clearly the simplicity and clarity that results from describing the simplex method explicitly as a sequence of matrix operations.

6.2

It should be emphasised that the dual solution $\mathbf{y}_0$ in the previous section refers to the primal problem as defined at the end of the first part of the two-part simplex method. In the first part, using the artificial objective function $z_1 + z_2 + \ldots + z_m$ (see section 4.3), the augmented matrix of coefficients of the constraint equations $(\tilde{\mathbf{A}},\tilde{\mathbf{b}})$ say, which does not contain $\mathbf{I}_m$, is converted to $(\mathbf{A},\mathbf{b})$ say which does contain $\mathbf{I}_m$. The conversion is effected by simplex stages each of which is equivalent to pre-multiplication by an elementary matrix $\mathbf{E}_k^*$ (see section 3.7), and the overall effect is that of pre-multiplication by the inverse of an $m \times m$ matrix $\mathbf{Q}$ say, which is the m columns of $\tilde{\mathbf{A}}$ which become $\mathbf{I}_m$ in $\mathbf{A}$. Thus

$$\mathbf{E}_k^* \ldots \mathbf{E}_2^* \mathbf{E}_1^* (\tilde{\mathbf{A}},\tilde{\mathbf{b}}) = (\mathbf{A},\mathbf{b}),$$
$$\mathbf{Q}^{-1}(\tilde{\mathbf{A}},\tilde{\mathbf{b}}) = (\mathbf{A},\mathbf{b}).$$

The matrix $\mathbf{Q}^{-1}$ will appear at the end of the first part in the m columns of the tableau corresponding to the artificial variables, and at that stage the first m columns are the matrix we have called $\mathbf{B}$.

Consider again the example from section 4.4. In our present notation

$$(\tilde{\mathbf{A}},\tilde{\mathbf{b}}) = \left(\begin{array}{ccccc|c} 1 & 1 & 1 & -1 & 0 & 2 \\ 1 & 2 & 1 & 0 & -1 & 5 \end{array}\right), \tag{6}$$

$$(\mathbf{A},\mathbf{b}) = \left(\begin{array}{ccccc|c} \frac{1}{2} & 1 & \frac{1}{2} & 0 & -\frac{1}{2} & \frac{5}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \frac{1}{2} \end{array}\right), \quad \mathbf{B} = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix}$$

(because x_1 and x_4 are basic variables in the optimum tableau), and

$$\mathbf{Q} = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix}, \quad \mathbf{Q}^{-1} = \begin{pmatrix} 0 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} = \mathbf{E}_2^* \mathbf{E}_1^*.$$

Denoting by $\tilde{\mathbf{y}}_0$ and $\mathbf{y}_0$ the dual solutions to the primal problem in the forms

minimise $\mathbf{c}^T\mathbf{x}$ *subject to* $\tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{b}}$, $\mathbf{x} \geq \mathbf{0}$ and

minimise $\mathbf{c}^T\mathbf{x}$ *subject to* $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ respectively,

it is easy to relate $\tilde{\mathbf{y}}_0$ and $\mathbf{y}_0$. Because if $\mathbf{y}_0$ satisfies

$$\mathbf{y}_0^T\mathbf{A} \leq \mathbf{c}^T, \ \mathbf{y}_0^T\mathbf{b} = g(\mathbf{y}_0) = f(\mathbf{x}_{opt}) = \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{b} \tag{7}$$

and $\tilde{\mathbf{y}}_0$ satisfies

$$\tilde{\mathbf{y}}_0^T\tilde{\mathbf{A}} \leq \mathbf{c}^T, \ \tilde{\mathbf{y}}_0^T\tilde{\mathbf{b}} = g(\tilde{\mathbf{y}}_0) = f(\mathbf{x}_{opt}) = \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{b}, \tag{8}$$

then substituting for $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ gives

$$\tilde{\mathbf{y}}_0^T(\mathbf{QA}) \leq \mathbf{c}^T \quad \text{and} \quad \tilde{\mathbf{y}}_0^T\mathbf{Qb} = \mathbf{c}_1^T\mathbf{B}^{-1}\mathbf{b}.$$

From (7) it is clear that

$$\tilde{\mathbf{y}}_0^T = \tilde{\mathbf{y}}_0^T\mathbf{Q}^{-1}.$$

If the columns of the simplex tableau in the first part were not discarded then they too would be subjected to pre-multiplication by $\mathbf{B}^{-1}$ during the second part and would provide $\tilde{\mathbf{y}}_0^T$ in the same way that the final tableau provides $\mathbf{y}_0$, because

$$\tilde{\mathbf{y}}_0^T = (\mathbf{c}_1^T\mathbf{B}^{-1})\mathbf{Q}^{-1} = \mathbf{c}_1^T(\mathbf{B}^{-1}\mathbf{Q}^{-1}).$$

In section 5.5 the solution of the dual problem as defined by the primal at the end of the first part of the example quoted above was $\mathbf{y}_0 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$. Hence the dual solution to the *l.p.p.* originally given in section 4.4 is

$$(4,2)\begin{pmatrix} 0 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{pmatrix} = (-2,3) = \tilde{\mathbf{y}}_0,$$

and we verify that $\tilde{\mathbf{y}}_0$ satisfies the dual constraints,

$$(-2,3)\begin{pmatrix} 1 & 1 & 1 & -1 & 0 \\ 1 & 2 & 1 & 0 & -1 \end{pmatrix} = (1,4,1,2,-3) \leq (1,5,2,2,7),$$

and has the same value as the primal optimum,

$$(-2,3)\begin{pmatrix} 2 \\ 5 \end{pmatrix} = 11.$$

Notice that the above analysis holds for any matrix $\mathbf{Q}$ not just the one identified by the first part of the two-part simplex method. Thus if a primal *l.p.p.* with constraint equations $\mathbf{Ax} = \mathbf{b}$ is converted to an equivalent *l.p.p.* by row operations on $(\mathbf{A},\mathbf{b})$ defined by $\mathbf{Q}$, then the new dual solution is the previous one multiplied by $\mathbf{Q}^{-1}$.

6.3 Theorems of Alternatives for Matrices

A number of interesting results which state that one of a pair of mutually exclusive possibilities concerning a general matrix $\mathbf{A}$ must be true, can be regarded as immediate consequences of the duality theorem.

In fact these theorems were established about fifty years before the duality theorem and part of one of them, the theorem of the separating hyperplane, is often used to provide a proof of the duality theorem which is independent of the simplex method. This is the approach used in the third proof of the simplex method; see Appendix 2.

Theorem 10. Farkas' Lemma, or the Theorem of the Separating Hyperplane

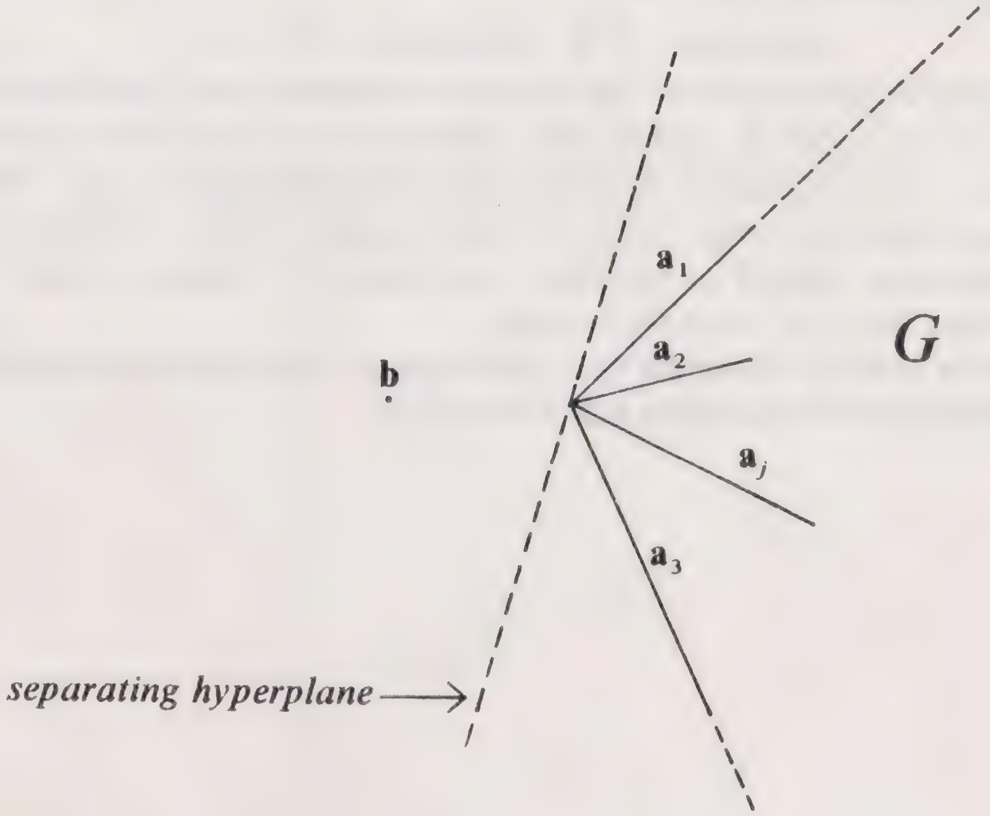
For any $m \times n$ matrix $\mathbf{A}$ and any m-vector $\mathbf{b}$

either (i) there is an n-vector $\mathbf{x}$ such that $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{Ax} = \mathbf{b}$,

or (ii) there is an m-vector $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{b} < 0$ and $\mathbf{y}^T\mathbf{A} \geq \mathbf{0}^T$ ∎

Strictly (i) false implies (ii) true is Farkas' Lemma, i.e. if there is no non-negative vector $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{b}$, then there is a vector $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{b} < 0$ and $\mathbf{y}^T\mathbf{A} \geq \mathbf{0}^T$. This is also called the theorem of the separating hyperplane because of the following interpretation:

> let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be the m-vector columns of $\mathbf{A}$ and let G be the set of all non-negative linear combinations of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. Thus $\mathbf{z} \in G$ if $\mathbf{z} = \mathbf{Ax}$ for $\mathbf{x} \geq \mathbf{0}$. The set G is convex (*ER*) and is in fact a *convex cone*. (A set G is a *cone* if $\mathbf{z} \in G$ implies $\alpha\mathbf{z} \in G$ for all $\alpha \geq 0$.) The vertex of the cone is the origin and corresponds to $\mathbf{x} = \mathbf{0}$.



The theorem says that any point $\mathbf{b}$ either belongs to G or can be separated from G by a hyperplane. The hyperplane is defined by $\mathbf{y}$ and consists of all points $\mathbf{z}$ such that $\mathbf{y}^T\mathbf{z} = 0$, and it separates

$\mathbf{b}$ and G because the projection of $\mathbf{b}$ onto $\mathbf{y}$, $\mathbf{y}^T\mathbf{b}$, is negative whereas the projection of any point of G onto $\mathbf{y}$ is positive, $\mathbf{y}^T\mathbf{A} \geq \mathbf{0}^T$, i.e. $\mathbf{y}^T\mathbf{a}_j \geq 0, j = 1, 2, n$.

The theorem may be proved as follows:

1. (i) true implies (ii) false.
   Let $\mathbf{x} \geq \mathbf{0}$ satisfy $\mathbf{Ax} = \mathbf{b}$ and consider any $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{A} \geq \mathbf{0}^T$. Then $(\mathbf{y}^T\mathbf{A})\mathbf{x} \geq \mathbf{0}$ and therefore $\mathbf{y}^T\mathbf{b} \geq 0$, because $\mathbf{Ax} = \mathbf{b}$, and therefore (ii) is false.
2. (ii) false implies (i) true.
   Note that it is not sufficient now to prove (ii) true implies (i) false because this leaves the possibility that both (i) and (ii) are false.

If (ii) is false then there is not a vector $\mathbf{y}$ such that

$$\mathbf{y}^T(-\mathbf{A}) \leq \mathbf{0}^T \quad \text{and} \quad \mathbf{y}^T(-\mathbf{b}) > 0. \tag{1}$$

The canonical dual *l.p.p.* is

$$\textit{maximise} \quad \mathbf{y}^T\mathbf{b} \quad \textit{subject to} \quad \mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T$$

and in this case we have $\mathbf{c} = \mathbf{0}$. Now $\mathbf{y} = \mathbf{0}$ satisfies the dual constraints and $\mathbf{y}^T\mathbf{b}$ has value 0, and by the assertion (1) $\mathbf{y} = \mathbf{0}$ is the optimum solution. By the duality theorem the canonical primal *l.p.p.* has an optimum solution (with value 0, which agrees with $\mathbf{c} = \mathbf{0}$) and thus the canonical primal is feasible, i.e. there is a vector $\mathbf{x}$ such that $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{Ax} = \mathbf{b}$. Thus (i) is true■

Similar results, including two theorems of alternatives for matrices, are mentioned in exercises 6.5, 6.6 and 6.7.

Exercises 6

1. Obtain the solution of the dual problems of the *l.p.p.s* of exercises 4.1, 4.3 (ii).
2. The *l.p.p.*

minimise $-2x_1 - 3x_2 - x_4$ *subject to* $x_1, x_2, x_3, x_4 \geq 0$ *and*

$$\begin{pmatrix} 0 & 3 & 1 & 0 \\ 1 & -2 & 4 & 0 \\ 1 & 4 & 2 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$$

is solved below. Obtain and check the optimum solution of the dual *l.p.p.*

0	3	1	0	1	0	2	$\frac{2}{1}$	
1	-2	4	0	0	1	2	$\frac{2}{4}$	←
1	4	2	1	0	0	4	$\frac{4}{2}$	
0	0	0	0	1	1	0		
-1	-1	-5	0	0	0	-4		
		↑						

$-\frac{1}{4}$	$\frac{7}{2}$	0	0	1	$-\frac{1}{4}$	$\frac{3}{2}$	$\frac{3}{7}$	←
$\frac{1}{4}$	$-\frac{1}{2}$	1	0	0	$\frac{1}{4}$	$\frac{1}{2}$		
$\frac{1}{2}$	5	0	1	0	$-\frac{1}{2}$	3	$\frac{3}{5}$	
$\frac{1}{4}$	$-\frac{7}{2}$	0	0	0	$\frac{5}{4}$	$-\frac{3}{2}$		
	↑							

$-\frac{1}{14}$	1	0	0	$\frac{2}{7}$	$-\frac{1}{14}$	$\frac{3}{7}$		
$\frac{3}{14}$	0	1	0	$\frac{1}{7}$	$\frac{3}{14}$	$\frac{5}{7}$	$\frac{10}{3}$	
$\frac{12}{14}$	0	0	1	$-\frac{10}{7}$	$-\frac{1}{7}$	$\frac{6}{7}$	1	←
0	0	0	0	1	1	0		
-2	-3	0	-1	0	0	0		
$-\frac{19}{14}$	0	0	0			$\frac{15}{7}$		
↑								

0	1	0	$\frac{1}{12}$	$\frac{1}{6}$	$-\frac{1}{12}$	$\frac{1}{2}$
0	0	1	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
1	0	0	$\frac{14}{12}$	$-\frac{5}{3}$	$-\frac{1}{6}$	1
0	0	0	$\frac{19}{12}$			$\frac{7}{2}$

$$\mathbf{x}^T_{opt} = (1, \tfrac{1}{2}, \tfrac{1}{2}, 0),\ f(\mathbf{x}_{opt}) = -\tfrac{7}{2}.$$

3. Suppose the *l.p.p.* in canonical primal form

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0},$$

where $\mathbf{A}$ is $m \times n$, has solution $\mathbf{x}_0$ and suppose that the dual problem has solution $\mathbf{y}_0$. The j-th equality constraint of the primal problem is now replaced by the (*j-th constraint*) + ($\lambda \times$ *the i-th constraint*). What is the solution of the new dual problem?

4. For the case $\mathbf{A} = \begin{pmatrix} \frac{1}{2} & 2 & 1 \\ 3 & \frac{1}{2} & 1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ the assertion (ii) of theorem 10 holds. Draw a diagram of the situation in the (y_1, y_2) plane and by inspection find the equation of a separating hyperplane (a line in this case). Explain how the equation of a separating hyperplane could be found in general.

5. One theorem of alternatives for matrices states that for any $m \times n$ matrix $\mathbf{A}$ and any m-vector $\mathbf{b}$

either (i) there is an n-vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$,

or (ii) there is an m-vector $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{A} \geq \mathbf{0}^T$, $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} < 0$.

Prove this theorem using the duality theorem. (Hint: connect the relationships in (i) and (ii) with the standard dual and standard primal *l.p.p.s* respectively.)

6. Use the duality theorem to establish the following theorem: ***Gordon's Theorem:*** For any $m \times n$ matrix $\mathbf{A}$

either (i) there is an n-vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$, $\mathbf{x} \geq \mathbf{0}$, $\mathbf{x} \neq \mathbf{0}$,

or (ii) there is an m-vector $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{A} > \mathbf{0}^T$.

(Hint: (i) is equivalent to:
there is an n-vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$, $\mathbf{x} \geq \mathbf{0}$, $\mathbf{e}^T\mathbf{x} = 1$, where $\mathbf{e}^T = (1, 1, \ldots, 1)$.)

7. Another theorem of alternatives for matrices: prove that for any $m \times n$ matrix $\mathbf{A}$ and any m-vector $\mathbf{b}$,

either there is an n-vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$,

or there is an m-vector $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{A} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq \mathbf{0}$.

8. Establish a stronger version of the equilibrium theorem for canonical form (see exercise 5.8), namely that dual constraints corresponding to primal variables basic at optimality are satisfied as equalities by the optimum solution of the dual.

9. Show that the relationship $\mathbf{c}^T\mathbf{x} = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{b}$ (see section 6.1) holds not just at the optimum stage, but at any stage of the simplex method, with $\tilde{\mathbf{c}}$ as defined in section 5.4, and $\mathbf{B}$ the $m \times m$ matrix consisting of the columns of $\mathbf{A}$ corresponding to the basic variables of $\mathbf{x}$ the current *b.f.s.*

NOTES

CHAPTER 7

THE REVISED SIMPLEX METHOD

7.1

As we have described it in section 3.2 the simplex method involves replacing a current system of equality constraints $\mathbf{A}'\mathbf{x} = \mathbf{b}'$ by an equivalent system $\mathbf{A}^*\mathbf{x} = \mathbf{b}^*$ at each stage and finding a corresponding new set of *e.c.c.s.* Nowadays this is very rarely what is done in practice. Two considerations motivate us to review our interpretation of the method and hence to produce a revised scheme for organising the calculations at each stage. The first is that we are really only identifying and solving an $m \times m$ system of equations. In doing so some (possibly most) of the variables will remain non-basic throughout the process, so that the arithmetic operations performed on the elements of the corresponding columns are, in a sense, unnecessary. The second is that when solving an $m \times m$ system of equations on a computer we know that the arithmetic operations are not performed exactly and that special methods should be used to minimise the effects of the arithmetic errors. One recommended method, called *Gaussian elimination with interchanges,* is described in Appendix 3 together with implications for the (revised) simplex method. In this chapter we concentrate on the first aspect and, since both parts of the two-part simplex method consist of solving by the simplex method a *l.p.p.* in which $\mathbf{A} \supset \mathbf{I}_m$, we can take as our starting point the *l.p.p.*

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0},$$

where $\mathbf{A} \supset \mathbf{I}_m$.

As we saw in section 3.7 each stage of the simplex method consists of premultiplying the current system $(\mathbf{A}',\mathbf{b}')$ by $\mathbf{E}_k^*$ to get $(\mathbf{A}^*,\mathbf{b}^*)$. Thus, at the end of the $(k-1)$-th stage we have

$$(\mathbf{A}',\mathbf{b}') = \mathbf{E}_{k-1}^* \mathbf{E}_{k-2}^* \ldots \mathbf{E}_2^* \mathbf{E}_1^* (\mathbf{A},\mathbf{b}), \tag{1}$$

and *e.c.c.s*

$$\mathbf{c}'^T = \mathbf{c}^T - \tilde{\mathbf{c}}^T(\mathbf{E}_{k-1}^* \ldots \mathbf{E}_2^* \mathbf{E}_1^*)\mathbf{A}, \tag{2}$$

where $(\tilde{\mathbf{c}})_i = c_{j_i}$ and the j_i-th column of $\mathbf{A}'$ is the i-th column of $\mathbf{I}_m$ (see section 5.4). If we denote by $\mathbf{B}$ the $m \times m$ matrix whose

i-th column is the j_i-th column of $\mathbf{A}$ (consistent with our notation in sections 5.4 and 6.1), then $\mathbf{E}^*_{k-1} \ldots \mathbf{E}^*_2 \mathbf{E}^*_1 = \mathbf{B}^{-1}$, and the three sets of coefficients which define the *l.p.p.*, currently $\mathbf{A}'$, $\mathbf{b}'$, $\mathbf{c}'$, are given by

$$\mathbf{A}' = \mathbf{B}^{-1}\mathbf{A}, \ \mathbf{b}' = \mathbf{B}^{-1}\mathbf{b} \quad \text{and} \quad \mathbf{c}'^T = \mathbf{c}^T - \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A}. \tag{3}$$

The next stage can be described as follows:

(i) Find $\min\limits_j c'_j = c'_t = \min\limits_j(\mathbf{c}^T - \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A})_j$
(if $\mathbf{c}' \geq \mathbf{0}$ we have the optimum solution).

(ii) Find $\min\limits_{\substack{i=1,2,\ldots,m \\ a'_{it} > 0}} \dfrac{b'_i}{a'_{it}} = \dfrac{b'_s}{a'_{st}} = \left(\dfrac{b'_i}{a'_{it}}\right) \min\limits_{\substack{i=1,\ldots,m \\ (\mathbf{B}^{-1}\mathbf{a}'_{*t})_i > 0}} (\mathbf{B}^{-1}\mathbf{b})_i/(\mathbf{B}^{-1}\mathbf{a}_{*t})_i$.

(iii) With s and t defined, evaluate $\mathbf{E}^*_k$.

(iv) Replace $\mathbf{B}^{-1}$ by $\mathbf{E}^*_k\mathbf{B}^{-1}$.

(v) Replace the s-th element of $\tilde{\mathbf{c}}$ by c_t, and j_s by t.

We now have the situation described by (3) again, with $\mathbf{A}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$ defined by the same equations but using the new $\mathbf{B}^{-1}$ and $\tilde{\mathbf{c}}$, so we can repeat steps 1, 2, 3, 4, 5 until the optimality criterion is satisfied. Remember that the actual matrix $\mathbf{B}^{-1}$ is stored, so we do not have to calculate the inverse of the matrix $\mathbf{B}$.

This approach, in which we use the original coefficients in $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{c}$ together with $\tilde{\mathbf{c}}$ and the matrix $\mathbf{B}^{-1}$ instead of the equivalent coefficients in $\mathbf{A}'$, $\mathbf{b}'$ and $\mathbf{c}'$, is called the **Revised Simplex Method.**

The precise implementation in practice has various alternatives, some of which are discussed in section 7.3.

7.2

It seems to be the custom to rejoice at this point at having found such an efficient improvement over the simplex tableau approach. Instead of calculating a completely new tableau we only have to calculate the new *e.c.c.s*, the new values of the current basic variables, $\mathbf{B}^{-1}\mathbf{b}$, the column vector $\mathbf{B}^{-1}\mathbf{a}_{*t}$ and the new $\mathbf{B}^{-1}$, as in steps 1, 2, 4 in the previous section. The celebrations are quite misguided, because the revised simplex method is no faster than the tableau approach for *general* matrices $\mathbf{A}$.

In both cases we can take account of the fact that $\mathbf{I}_m$ is present and so regard the tableaux $\mathbf{A}$ and $\mathbf{A}'$ as $m \times (n-m)$ matrices. What

is usually overlooked is that evaluating $\tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A}$ involves the same number of arithmetic operations as forming $\mathbf{A}^*$ from $\mathbf{A}'$. *Essentially,* if we put $\tilde{\mathbf{c}}^T\mathbf{B}^{-1} = \mathbf{d}^T$ then evaluating $\mathbf{d}^T\mathbf{A}$ requires $m(n-m)$ multiplications and $(m-1)(n-m)$ additions, whereas evaluating $\mathbf{A}^*$ (see ① and ② of section 3.2) requires one division, $m(n-m)$ multiplications and $(m-1)(n-m)$ additions.

Basically both approaches require $m(n-m)$ additions and multiplications for the main part of the calculations at each stage. However, the revised simplex method, in one or other of its implementations, is now the standard method for solving *l.p.p.s* in practice, so we need to see why.

The crucial point is that the revised simplex method always involves the original matrix $\mathbf{A}$, instead of a sequence of changing matrices $\mathbf{A}'$. For most large problems in practice $\mathbf{A}$ is *sparse,* that is most of the elements a_{ij} are zero (less than 20% non-zero is not uncommon), and of course all the arithmetic operations involving addition of zero or multiplication by zero can be omitted. Remember that in theory replacing a_{ij} by $a_{ij} + 0$, or multiplying 0 by a_{sj} is the same as omitting the operation, whereas in practice many (perhaps most) computers take as long to add zero or multiply by zero as they do to add or multiply by any other number.

The tableau operations tend to *fill-in* the zero elements so that $\mathbf{A}'$ becomes less and less sparse, and so even if we were to avoid actually performing operations with zeros there are more operations to perform at each stage. On the other hand, the revised simplex operations are to evaluate $\mathbf{d}^T = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}$, where neither of $\tilde{\mathbf{c}}^T$ and $\mathbf{B}^{-1}$ is to be regarded as sparse (although $\mathbf{B}^{-1}$ certainly is initially), and then to evaluate $\mathbf{d}^T\mathbf{A}$. When $\mathbf{A}$ is sparse, many *empty* operations (involving zeros) can be omitted, and exactly the same operations at every stage.

This then is the reason for the greater efficiency of the revised simplex method, but the savings are non-existent if $\mathbf{A}$ is not significantly sparse, and are not realised if we do not take advantage of the sparseness. This suggests, correctly, that a good revised simplex computer program is quite a complicated affair. In practice, a sparse $\mathbf{A}$ is not stored as an $m \times n$ matrix at all; instead only the non-zero elements together with their row and column indices, a_{ij}, i, j, are stored and the arithmetic operations are organised in terms of this information.

7.3

There are three distinct ways of implementing the revised simplex method, although in practice different aspects of each can be combined. In the brief discussion of each that follows, we assume that advantage will be taken of the sparseness.

(i) The implementation can be explicitly as described in section 7.1, with $\mathbf{B}^{-1}$ stored as an $m \times m$ matrix and "updated" at each stage by pre-multiplication by $\mathbf{E}_k^*$, and $\mathbf{b}' = \mathbf{B}^{-1}\mathbf{b}$, $\mathbf{d}^T = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}$ and $\mathbf{d}^T\mathbf{A}$ evaluated as vector $\times$ matrix products. Note that $\mathbf{E}_k^*$ has a very simple form and $\mathbf{E}_k^*\mathbf{B}^{-1}$ would *not* be evaluated as a general matrix product but rather as a sequence of row operations on $\mathbf{B}^{-1}$. Also, we would probably have an m-vector $\mathbf{j}$, whose elements $j_1, j_2, \ldots, j_m$ are the column indices of the columns of $\mathbf{I}_m$, so that

$$d_k = (\tilde{\mathbf{c}}^T\mathbf{B}^{-1})_k, \ k = 1, 2, \ldots, m,$$

is given by $\Sigma_{i=1}^m c_{j_i} (\mathbf{B}^{-1})_{ik}$.

(ii) Instead of storing $\mathbf{B}^{-1}$ explicitly as an $m \times m$ matrix, we can store it implicitly in product form, because $\mathbf{E}_k^*$ is obtained at each stage and

$$\mathbf{B}^{-1} = \mathbf{E}_k^* \mathbf{E}_{k-1}^* \ldots \mathbf{E}_2^* \mathbf{E}_1^* .$$

Then an expression such as $\mathbf{B}^{-1}\mathbf{b}$ can be evaluated by evaluating successively $\mathbf{E}_1^*\mathbf{b}$, $\mathbf{E}_2^*(\mathbf{E}_1^*\mathbf{b})$, $\mathbf{E}_3^*(\mathbf{E}_2^*(\mathbf{E}_1^*\mathbf{b}))$ etc. Remember that each of these products can be evaluated efficiently by taking into account the special form of $\mathbf{E}_k^*$, and that to store $\mathbf{E}_k^*$ we only need to store the single non-trivial column of $\mathbf{E}_k^*$ together with the corresponding integer column index. Thus $\mathbf{E}_1^*, \mathbf{E}_2^*, \ldots, \mathbf{E}_m^*$ can be stored in the same amount of space that $\mathbf{B}^{-1}$ requires. More stages would require extra storage space, or a compromise between the two approaches. The advantage of this approach is that each column vector representing an $\mathbf{E}_k^*$ will be sparse if $\mathbf{A}$ is sparse and this can be used to save storage and to reduce the number of operations performed.

(iii) The third approach brings us back full circle to the observation made near the end of section 2.9, that solving a *l.p.p.* really only requires solving an $m \times m$ system of linear equations.
The three equations involving $\mathbf{B}^{-1}$ at each stage are

$$\mathbf{d}^T = \tilde{\mathbf{c}}^T\mathbf{B}^{-1}, \ \mathbf{b}' = \mathbf{B}^{-1}\mathbf{b} \quad \text{and} \quad \mathbf{a}'_{*t} = \mathbf{B}^{-1}\mathbf{a}_{*t}. \tag{1}$$

These can be regarded as three $m \times m$ systems of linear equations for the unknown vectors $\mathbf{d}^T$, $\mathbf{b}'$ and $\mathbf{a}'_{*t}$ which all involve the same matrix of coefficients $\mathbf{B}$. The matrix $\mathbf{B}$ consists of the m columns

of $\mathbf{A}$ identified by the integer elements of the vector $\mathbf{j}$ of the previous section. Thus instead of storing $\mathbf{B}^{-1}$, updating it at each stage and explicitly forming the products (1), we can solve the three systems of equations

$$\mathbf{B}\mathbf{a}'_{*t} = \mathbf{a}_{*t}; \quad \mathbf{B}\mathbf{b}' = \mathbf{b} \quad \text{and} \quad \mathbf{B}^T\mathbf{d} = \tilde{\mathbf{c}}. \tag{2}$$

This would appear to be very inefficient, but the three systems can be solved with little more effort than is needed to solve one of them (see Appendix 3), and it is possible to update information from the previous stage to avoid most of the calculations required (see {12}, chapter 1.2 of {8} and also exercise 7.3). The important aspect of this approach is that a method of solving the equations that is known to produce satisfactorily accurate solutions can be used. In addition, one can periodically revert to the equations (2) and solve them without reference to earlier stages, to prevent successive arithmetic errors building up.

This approach ensures that the simplex method is numerically stable, and should be the standard approach in practice.

Exercises 7

1. Assuming that **A** is not sparse, evaluate precisely the number of arithmetic operations (additions, multiplications and divisions) needed to perform one stage of the simplex method
   a) using the tableau,
   b) using the 7.3(i) implementation of the revised simplex method.
2. Re-solve, using implementations 7.3(i) and 7.3(ii) of the revised simplex method, a *l.p.p.* previously solved using the simplex tableau, e.g. the problem in section 3.4, the problem in section 4.4, exercise 4.1.
3. Suppose an $n \times n$ matrix $\mathbf{B}_a$ is obtained by replacing the s-th column of an $n \times n$ matrix $\mathbf{B}$ by an n-vector $\mathbf{a}$. Explain how $\mathbf{B}_a^{-1}$ may be obtained efficiently if $\mathbf{B}^{-1}$ is available. (See step 4, section 7.1 and section 7.3.)

4. With particular reference to the multiplication details of p 92 (i) and p. 31 ①, ② and ③ explain/devise a simple variant of the simple/revised simplex method (as we have described it) which would make the revised simplex method more efficient in general. (Hint: Section 3.3)

NOTES

NOTES

CHAPTER 8

PARAMETRIC LINEAR PROGRAMMING AND SENSITIVITY ANALYSIS

8.1

If a *l.p.p.* is solved, and then a small change is made, such as one coefficient a_{ij}, b_i, or c_j changed, or one constraint removed, one would hope that the solution of the new problem could be obtained without having to start all over again. For certain changes this is the case and so the effects of, say, a changing price or a change in resources can be determined efficiently. Doing so effectively determines the sensitivity of the solution to the particular coefficient or constraint involved. An introduction to this aspect of linear programming is given in the sections 8.3 to 8.6, by considering four particular changes. This section and the following section are concerned with a similar aspect in which the objective function $f(\mathbf{x})$ depends linearly on a parameter λ, and we require the optimum solution as a function of λ. This is usually called *parametric linear programming,* although this term would also be appropriate if a parameter were present in $\mathbf{A}$ or $\mathbf{b}$.

The vector of cost coefficients may be denoted by $\mathbf{c}(\lambda) = \mathbf{c} + \lambda\mathbf{d}$, so that the objective function $f = f(\mathbf{x},\lambda) = (\mathbf{c} + \lambda\mathbf{d})^T\mathbf{x} = \mathbf{c}^T\mathbf{x} + \lambda\mathbf{d}^T\mathbf{x}$.

Suppose that the interval $\lambda_L \leq \lambda \leq \lambda_U$ is of interest, and as usual assume $\mathbf{A} \supset \mathbf{I}_m$. The *l.p.p.* with $\lambda = \lambda_L$ may be solved in the normal way, but in terms of the tableau approach, we can replace the c-row of *e.c.c.s* by two rows, a c-row and a d-row, which are initially $\mathbf{c}^T$ and $\mathbf{d}^T$ respectively. The *e.c.c.s* at any stage are given by $\mathbf{c}'(\lambda) = \mathbf{c}' + \lambda_L\mathbf{d}'$, where $\mathbf{c}'$ and $\mathbf{d}'$ at every stage are each obtained by an appropriate row operation. This presents no difficulties; the usual c_j' becomes $c_j' + \lambda_L d_j'$ and so the optimality criterion is

$$c_j'(\lambda) = c_j' + \lambda_L d_j' \geq 0, j = 1, 2, \ldots, n.$$

There are two possibilities:

(i) for $\lambda = \lambda_L$ an optimum solution is obtained, and

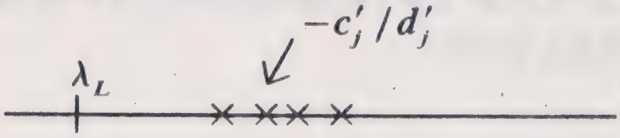
(ii) for $\lambda = \lambda_L$ and some t, at some stage we find

$$c_t'(\lambda) < 0 \quad \text{and} \quad a_{it}' \leq 0, \; i = 1, 2, \ldots, m.$$

If (i) is the case, we would expect that in general λ can be moved from λ_L without violating the conditions

$$c'_j(\lambda) \geq 0, \; j = 1, 2, \ldots, n, \tag{1}$$

i.e. for some range of λ, $\lambda_- \leq \lambda \leq \lambda_+$, $c'_j(\lambda) \geq 0$, where $\lambda_- \leq \lambda_L \leq \lambda_+$. For $d'_j < 0$ we know that $\lambda_L \leq -c'_j/d'_j$, and we require $\lambda \leq -c_j/d_j$.



For $d'_j > 0$ we know that $\lambda_L \geq -c'_j/d'_j$, and we require $\lambda \geq -c'_j/d'_j$.

Thus the inequalities (1) are satisfied by $\lambda_- \leq \lambda \leq \lambda_+$,

$$\text{where } \lambda_- = \max_{\substack{j=1,2,\ldots,n \\ d'_j > 0}} -c'_j/d_j,$$

$$\lambda_+ = \min_{\substack{j=1,2,\ldots,n \\ d'_j < 0}} -c'_j/d'_j,$$

and λ_-, λ_+ are $-\infty$, $+\infty$ if all d'_j are ≤ 0, ≥ 0 respectively.

The current optimum solution $\mathbf{x}_0$ remains the optimum solution for $\lambda_- \leq \lambda \leq \lambda_+$; $\lambda_- \leq \lambda_L$ and the value of λ_- may or may not be of interest. If $\lambda_+ = +\infty$ or $\lambda_+ \geq \lambda_U$ the parametric *l.p.p.* is solved, and there is a single optimum solution for the whole of the range of interest of λ. Note that although the optimum solution $\mathbf{x}_0$ does not change, the value of $f(\mathbf{x}_0, \lambda)$ varies linearly with λ.

If, however, λ_+ is finite and $\lambda_+ < \lambda_U$, then we must have $\lambda_+ = -c'_t/d'_t$ for some t, $1 \leq t \leq n$, and so for $\lambda > \lambda_+$, $c'_t(\lambda) < 0$, and if $a'_{it} \leq 0$, $i = 1, 2, \ldots, m$, the *l.p.p.* with $\lambda > \lambda_+$ has feasible solutions whose values are unbounded below. Otherwise $a'_{it} > 0$ for some i, $1 \leq i \leq n$, and if we perform the pivotal operations of the simplex method with a'_{st} as pivot we obtain an optimum simplex tableau (for $\lambda = \lambda_+$) in which x_t is a basic variable. This returns us to the beginning of case (i) with λ_+ instead of λ_L, so we put $\lambda_+ = \lambda_1$ say and repeat the procedure. The next time we find $\lambda_+ \geq \lambda_U$, or feasible solutions with values unbounded below or we find a new λ_+, say λ_2. Thus we generate a sequence of *characteristic values* $\{\lambda_i\}$, $\lambda_L \leq \lambda_1 \leq \lambda_2 \ldots \leq \lambda_k$, $\lambda_k \geq \lambda_U$. It may happen that $\lambda_i = \lambda_{i+1}$, but it can be shown that the set of basic variables (and generally the optimum solutions) corresponding to λ_i and λ_{i+1} are different, and cannot occur again.

If (ii) is the case, either for $\lambda = \lambda_L$ or for any $\lambda = \lambda_i$, then the *l.p.p.* for this value of λ (we shall use $\lambda = \lambda_L$ for convenience) has feasible solutions but no optimum solutions. Suppose that $c'_i(\lambda_L) < 0$, i.e. $c'_i + \lambda_L d'_i < 0$, then if $d'_i \leq 0$, $c'_i(\lambda) < 0$ for any $\lambda \geq \lambda_L$, and with $a'_{it} \leq 0$, $i = 1, 2, \ldots, m$, we conclude that there is no optimum solution for any $\lambda \geq \lambda_L$. On the other hand, if $d'_i > 0$, then $c'_i(\lambda) < 0$ for $\lambda_L < \lambda < -c'_i/d'_i$.

So consider $\lambda = \lambda' = -c'_i/d'_i$. If $\lambda \geq \lambda_U$ there are feasible solutions with unbounded values for $\lambda_L \leq \lambda \leq \lambda_U$. Assuming that $\lambda' < \lambda_U$ we essentially just return to the simplex method. If $c'_j(\lambda') \geq 0, j = 1, 2, \ldots, n$, we have an optimum feasible solution, so we have case (i) with λ' instead of λ_L. If $c'_j(\lambda') < 0$ for some j, we continue the simplex method, and with $\lambda = \lambda'$ we arrive again at one of the two possibilities (i) and (ii), and we continue until either $\lambda' \geq \lambda_U$ or $\lambda_k \geq \lambda_U$.

The procedure in practice is simpler than the above analysis suggests, as the example in section 8.2 demonstrates. We observe that the situation at a characteristic value is essentially that with which exercise 3.6 is concerned.

We also observe that if a *l.p.p.* is solved and then $\mathbf{c}$ is changed to $\mathbf{c}_0$ say, it is a simple matter to solve the modified problem. We just replace $\mathbf{c}'_{opt}$ by $\mathbf{c}_0$, convert to equivalent cost coefficients $\mathbf{c}'$ by the operations described in section 4.2, and if $\mathbf{c}'_j < 0$ for any j we just continue with the simplex method. If $\mathbf{c}' \geq \mathbf{0}$ then the optimum solution is unchanged, but the optimum value changes from $\mathbf{c}^T\mathbf{x}_{opt}$ to $\mathbf{c}_0^T\mathbf{x}_{opt}$. We consider one aspect of this in more detail in section 8.6.

8.2 Example

Solve the *l.p.p.*

minimise $\{(-1,-2,-1,0,0,0) + \lambda(1,0,-3,2,0,-6)\}\mathbf{x}$

subject to $\mathbf{x} \geq \mathbf{0}$ *and*

$$\begin{pmatrix} 2 & 1 & -1 & 1 & 0 & 0 \\ 2 & -1 & 5 & 0 & 1 & 0 \\ 4 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 6 \\ 6 \end{pmatrix}, \tag{1}$$

for $0 \leq \lambda \leq \infty$.

This is the problem of section 3.4 with $\mathbf{c}$ replaced by $\mathbf{c} + \lambda\mathbf{d}$, where $\mathbf{d}^T = (1,0,-3,2,0,-6)$, and with $\lambda_L = 0$, $\lambda_U = \infty$.

Solving the *l.p.p.* with $\lambda = \lambda_L = 0$ is the same as solving the *l.p.p.* of section 3.4 so we can start with the optimum tableau of that section,

just adding a *d-row*, and converting to the appropriate form with $d_2' = d_3' = d_6' = 0$ because x_2, x_3 and x_6 were the basic variables at optimality.

Thus the first stage in this case is given by

$$\begin{array}{cccccc|c|c}
3 & 1 & 0 & \frac{5}{4} & \frac{1}{4} & 0 & 4 & 16 \\
1 & 0 & 1 & \frac{1}{4} & \frac{1}{4} & 0 & 2 & 8 \leftarrow \\
0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & 0 & \\
\hline
6 & 0 & 0 & \frac{11}{4} & \frac{3}{4} & 0 & 10 & \\
1 & 0 & -3 & 2 & 0 & -6 & 0 & \\
4 & 0 & 0 & -\frac{25}{4} & -\frac{9}{4} & 0 & 6 & \\
 & & & & \uparrow & & &
\end{array} \tag{2}$$

Here we have case (i), so

$\lambda_- = max\,\{-\frac{6}{4}\} = -\frac{3}{2}$,

$\lambda_+ = min\,\{\frac{11}{25}, \frac{3}{9}\} = \frac{1}{3} = \lambda_1,\ t = 5,$

and hence via the θ-column $s = 2$. Pivoting on a_{25}' leads to the next tableau.

$$\begin{array}{rcccccc|c|c}
 & 2 & 1 & -1 & 1 & 0 & 0 & 2 & 2 \leftarrow \\
 & 4 & 0 & 4 & 1 & 1 & 0 & 8 & 8 \\
 & 2 & 0 & 2 & -1 & 0 & 1 & 4 & \\
\hline
 & 3 & 0 & -3 & 2 & 0 & 0 & 4 & \\
 & 13 & 0 & 9 & -4 & 0 & 0 & 24 & \\
 & & & & \uparrow & & & & \\
\mathbf{c}'(\lambda_1)^T & \frac{22}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 12 &
\end{array} \tag{3}$$

Thus $\mathbf{x}_{opt} = (0,2,0,0,8,4)^T$ when $\lambda = \lambda_1 = \frac{1}{3}$, and we have added the row $\mathbf{c}'(\lambda)$ for $\lambda = \lambda_1$ to confirm that $\mathbf{c}'(\lambda_1) \geq 0$. For $\lambda = \lambda_1$ the optimum value is -12. This is again case (i) so

$\lambda_- = max\,\{-\frac{3}{13}, \frac{3}{9}\} = \frac{1}{3}$ (= λ_1 of course),

$\lambda_+ = min\,\{+\frac{2}{4}\} = \frac{1}{2} = \lambda_2,\ t = 4,\ and\quad s = 1.$

Thus we obtain

$$\begin{array}{rcccccc|c|}
 & 2 & 1 & -1 & 1 & 0 & 0 & 2 \\
 & 2 & -1 & 5 & 0 & 1 & 0 & 6 \\
 & 4 & 1 & 1 & 0 & 0 & 1 & 6 \\
\hline
 & -1 & -2 & -1 & 0 & 0 & 0 & 0 \\
 & 21 & 4 & 5 & 0 & 0 & 0 & 32 \\
\mathbf{c}'(\lambda_2)^T & \frac{19}{2} & 0 & \frac{3}{2} & 0 & 0 & 0 & 16
\end{array} \tag{4}$$

Thus $\mathbf{x}_{opt} = (0,0,0,2,6,6)^T$ when $\lambda \geq \lambda_2 = \frac{1}{2}$, and we have again added $\mathbf{c}'(\lambda)^T$ to confirm that $\mathbf{c}'(\lambda_2) \geq \mathbf{0}$.

The characteristic values of the parameter λ are $\frac{1}{3}$ and $\frac{1}{2}$ and
for $0 \leq \lambda \leq \frac{1}{3}$ $(-\frac{3}{2} \leq \lambda \leq \frac{1}{3}$ in fact), $\mathbf{x}_{opt} = (0,4,2,0,0,0)^T = \mathbf{x}_0$ say,
for $\frac{1}{3} \leq \lambda \leq \frac{1}{2}$, $\mathbf{x}_{opt} = (0,2,0,0,8,4)^T = \mathbf{x}_1$, and
for $\frac{1}{2} \leq \lambda$, $\mathbf{x}_{opt} = (0,0,0,2,6,6)^T = \mathbf{x}_2$.

We observe that

$$f(\mathbf{x}) = (\mathbf{c}^T + \lambda\mathbf{d}^T)\mathbf{x} = -12 \quad \text{for} \quad \lambda = \tfrac{1}{3},\ \mathbf{x} = \mathbf{x}_0 \quad \text{and}$$
$$f(\mathbf{x}) = -12 \quad \text{for} \quad \lambda = \tfrac{1}{3},\ \mathbf{x} = \mathbf{x}_1.$$

We also observe that tableau (4) is the initial tableau for this problem from section 3.4: the two stages in this instance have just reversed the stages for the original version of the problem without the parameter λ.

8.3 Removal of a Constraint

Suppose that a *l.p.p.* is solved, and then the i-th (original) constraint is removed and we wish to know whether the optimum solution we already have is still optimum. Denote the two *l.p.p.s* by

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \tag{1}$$

and

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{b}},\ \mathbf{x} \geq \mathbf{0}, \tag{2}$$

where $(\mathbf{A},\mathbf{b})$ is $m \times (n+1)$ and $(\tilde{\mathbf{A}},\tilde{\mathbf{b}})$ is $(m-1) \times (n+1)$, and denote the optimum solution of (1) by $\mathbf{x}_0$.

We cannot simply remove the i-th constraint from the final tableau for (1) because in general all other rows have had a multiple of the i-th row added to them. The crucial question is whether the i-th constraint of (1) is active for $\mathbf{x} = \mathbf{x}_0$, because if it is not, then removing it will not alter the situation. Thus if (1) is derived from a problem with inequality constraints and the i-th slack or surplus variable is positive in $\mathbf{x}_0$ then $\mathbf{x}_0$ is optimum for (2). For *genuine* equality constraints in (1) all are active and if they are independent, removing any one changes R and we would expect that $\mathbf{x}_0$ would not still be optimum. We can be more precise if we examine the dual in conjunction with the equilibrium theorem. Let $\mathbf{y}_0$ be the optimum solution of the dual of (1), then $(\mathbf{y}_0)_i = 0$ implies that $\mathbf{x}_0$ is optimum for (2) (*ER*).

8.4 Introduction of a Further Constraint

Suppose that a *l.p.p.* is solved and then a further independent constraint is imposed. We use the same notation (1) and (2) for the

two *l.p.p.s* as in the previous section, where now $(\tilde{\mathbf{A}}, \tilde{\mathbf{b}})$ is $(m+1)\times(n+1)$ and

$$a_{m+1,1}x_1 + a_{m+1,2}x_2 + \dots + a_{m+1,n}x_n = b_{m+1} \tag{3}$$

denotes the extra constraint. If $\mathbf{x}_0$ satisfies (3) then it is the optimum solution for the new *l.p.p.* (*ER*).

If $\Sigma_{j=0}^{n} a_{m+1,j}(\mathbf{x}_0)_j \neq b_{m+1}$ then we may proceed as follows, assuming for convenience that $x_1, x_2, \dots, x_m$ are the basic variables in $\mathbf{x}_0$.

Insert the extra $(m+1)$-th row into the optimum tableau and subtract $a_{m+1,i} \times$ (i-th row) from this $(m+1)$-th row for $i = 1, 2, \dots, m$ to obtain $0, 0, \dots, 0, a'_{m+1,m+1}, a'_{m+1,m+2}, \dots, a'_{m+1,n}, b'_{m+1}$. If $b'_{m+1} < 0$ multiply this row by -1. We now have a *l.p.p.* in which the columns of the $(m+1)\times n$ matrix of coefficients include m columns of the $(m+1)\times(m+1)$ unit matrix $\mathbf{I}_m$, which we can solve by the two-part simplex method with only one artificial variable in the first part.

8.5 Variation of b

We consider only the change in which $\mathbf{b}$ is replaced by $\tilde{\mathbf{b}} = \mathbf{b} + \delta\mathbf{e}_k$, i.e. $\tilde{b}_i = b_i$, $i = 1, 2, \dots, k-1, k+1, \dots, m$; $\tilde{b}_k = b_k + \delta$. Denoting the optimum solution for the *l.p.p.* (1) of section 8.3 (i.e. $\delta = 0$) by $\mathbf{x}_0$, then

$$\mathbf{A}\mathbf{x}_0 \neq \mathbf{b} + \delta\mathbf{e}_k \quad \text{for} \quad \delta \neq 0,$$

so $\mathbf{x}_0$ cannot still be the optimum solution. However, the values of the basic variables in $\mathbf{x}_0$ are given by $\mathbf{B}^{-1}\mathbf{b}$ for the appropriate $\mathbf{B}^{-1}$, and if $\mathbf{B}^{-1}\tilde{\mathbf{b}}$ is non-negative the basic variables at optimality are unchanged and their values are given by $\mathbf{B}^{-1}\tilde{\mathbf{b}}$, because the corresponding *e.c.c.s* $\mathbf{c}'$ are still given by

$$\mathbf{c}' = \mathbf{c}^T - \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A}.$$

It is easy to find the range of δ for which the basic variables at optimality are unchanged;

$$\mathbf{B}^{-1}\tilde{\mathbf{b}} = \mathbf{B}^{-1}\mathbf{b} + \delta\mathbf{B}^{-1}\mathbf{e}_k = \mathbf{b}' + \delta\mathbf{b}^{(k)} \quad \text{say},$$

where $\mathbf{b}'$ denotes $\mathbf{B}^{-1}\mathbf{b}$, and $\mathbf{b}^{(k)}$ denotes the k-th column of $\mathbf{B}^{-1}$.

Thus we require $b'_i + \delta b_i^{(k)} \geq 0$, and so the set of basic variables at optimality is unchanged for

$$\max_{b_i^{(k)}>0} \frac{-b'_i}{b_i^{(k)}} \leq \delta \leq \min_{b_i^{(k)}<0} \frac{-b'_i}{b_i^{(k)}}.$$

For the situation in which $\mathbf{b}$ depends linearly on a parameter λ, i.e. $\mathbf{b}(\lambda) = \mathbf{b} + \lambda\tilde{\mathbf{b}}$ we can use the analysis of this section to obtain a sequence

of characteristic values of λ as in section 8.1, for each of which the optimum solution has a particular set of basic variables. Instead of doing so, we only observe that the same procedure as that developed in section 8.1 can be used if we work with the dual problem instead.

We also mention, without giving any details, that if a_{ij} is replaced by $a_{ij} + \delta$ then $\mathbf{B}^{-1}$ may or may not be changed, depending on whether x_i is basic or non-basic at optimality. Denoting the new $\mathbf{A}$ by $\tilde{\mathbf{A}}$ and the new $\mathbf{B}^{-1}$ by $\tilde{\mathbf{B}}^{-1}$ then the feasibility and optimality criteria,

$$\mathbf{b}' = \tilde{\mathbf{B}}^{-1}\mathbf{b} \geq \mathbf{0} \quad \text{and} \quad \mathbf{c}' = \mathbf{c} - \tilde{\mathbf{c}}^T\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{A}} \geq \mathbf{0}$$

can be used to determine the effects of the change (see {9}).

8.6 Variation of c

In this section we amplify the remarks at the end of section 8.1 and analyse the effects of a change in one cost coefficient.

Suppose c_k is changed to $c_k + \delta$ and the optimum solution of the *l.p.p.*

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}$$

is $\mathbf{x}_0$. The optimality criterion is

$$\mathbf{c}'^T = \mathbf{c}^T - \tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A} \geq \mathbf{0}^T$$

so we distinguish the two cases

(i) x_k non-basic at optimality, and

(ii) x_k basic at optimality, so that c_k appears in $\tilde{\mathbf{c}}^T$.

In case (i) $\tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A}$ is unchanged, so $\mathbf{x}_0$ is still an optimum solution and the optimum value $\mathbf{c}^T\mathbf{x}$ is unchanged provided that the new k-th *e.c.c.* is still non-negative,

$$\text{i.e.} \quad c'_k + \delta \geq 0 \quad \text{or} \quad \delta \geq -c'_k,$$

where c'_k is the k-th equivalent cost coefficient at optimality.

In case (ii), suppose that the k-th column of the final tableau is the s-th column of $\mathbf{I}_m$, and consider $c'_j(\delta)$, the j-th *e.c.c.* after the change of c_k to $c_k + \delta$, for $j = 1, 2, \ldots, n$.

If $j = k$,

$$\begin{aligned} c'_k(\delta) &= c_k + \delta - (\tilde{\mathbf{c}}^T + \delta\mathbf{e}_s^T)(\mathbf{B}^{-1}\mathbf{A})_{*k} \\ &= c_k + \delta - (\tilde{\mathbf{c}}^T + \delta\mathbf{e}_s^T)\mathbf{e}_s \\ &= c_k + \delta - (c_k + \delta) = 0, \end{aligned}$$

so $c'_k(\delta) \geq 0$ for any value of δ.

If $j \neq k$ and x_j is basic, then $c'_j(\delta) = c_j - (\tilde{\mathbf{c}}^T + \delta\mathbf{e}_s^T)(\mathbf{B}^{-1}\mathbf{A})_{*j}$, where $(\mathbf{B}^{-1}\mathbf{A})_{*j} = \mathbf{e}_{\ell}$ for $\ell \neq s$, so that $c'_j(\delta) = c_j - c_j = 0$ and $c'_j(\delta) \geq 0$ for any value of δ.

If $j \neq k$ and x_j is non-basic, then

$$c_j'(\delta) = c_j - (\tilde{\mathbf{c}}^T + \delta \mathbf{e}_s^T)(\mathbf{B}^{-1}\mathbf{A})_{*j}$$
$$= c_j - (\tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A})_j - \delta a_{sj}',$$

where a_{sj}' is the (s,j)-th element of $\mathbf{B}^{-1}\mathbf{A}$, and we know that

$$c_j - (\tilde{\mathbf{c}}^T\mathbf{B}^{-1}\mathbf{A})_j \quad \text{is the } j\text{-th } e.c.c. \ c_j' \text{ at optimality.}$$

Thus $c_j'(\delta) = c_j' - \delta a_{sj}'$, and

$$c_j'(\delta) \geq 0 \quad \text{if} \quad \begin{cases} \delta \leq c_j'/a_{sj}' & \text{for} \quad a_{sj}' > 0 \quad \text{or} \\ \delta \geq c_j'/a_{sj}' & \text{for} \quad a_{sj}' < 0, \end{cases}$$

and for $\mathbf{x}_0$ still to be the optimum solution we require

$$c_j'(\delta) \geq 0, \ j = 1, 2, \ldots, n.$$

Hence for any k, $1 \leq k \leq n$, $\mathbf{x}_0$ remains the optimum solution when c_k is changed to $c_k + \delta$ for

$$\max_{a_{sj}'<0} \frac{c_j'}{a_{sj}'} \leq \delta \leq \min_{a_{sj}'>0} \frac{c_j'}{a_{sj}'} \qquad \text{if} \quad x_k \text{ is basic and } (\mathbf{B}^{-1}\mathbf{A})_{*k} = \mathbf{e}_s$$

and x_j is non-basic

$$\text{or} \quad \delta \geq -c_k' \qquad \text{if} \quad x_k \text{ is non-basic.}$$

If x_k is non-basic the optimum value $\mathbf{c}^T\mathbf{x}_0$ is unchanged; if x_k is basic the optimum value increases by δx_k.

Exercises 8

1. Solve the *l.p.p.*

$$\textit{minimise} \quad (\mathbf{c} + \lambda\mathbf{d})^T\mathbf{x} \quad \textit{subject to}$$
$$\begin{pmatrix} 0 & 3 & 1 & 0 \\ 1 & -2 & 4 & 0 \\ 1 & 4 & 2 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}, \quad \mathbf{x} \geq \mathbf{0},$$

where $\mathbf{c}^T = (-2,-3,0,-1)$ and $\mathbf{d}^T = (0,1,0,-3)$. There is only one characteristic value; choose $\lambda_L = 0$ (see exercise 6.2). Check your result by considering $\lambda = \frac{1}{2}$ and $\lambda = 1$ and inserting the corresponding $\mathbf{c}(\lambda)$ in the appropriate tableau.

2. Solve the *l.p.p.*

$$\textit{maximise} \quad x_1 + x_2 \quad \textit{subject to} \quad x_1, x_2 \geq 0 \quad \textit{and}$$
$$\begin{aligned} 3x_1 + 2x_2 &\leq 6 \\ x_1 - x_2 &\leq 1 \\ 2x_1 + 3x_2 &\leq 6. \end{aligned}$$

Use the dual problem to determine which constraint may be omitted, without changing the optimum solution. Verify your results with a diagram.

3. For the *l.p.p.*

$$\textit{minimise} \quad (\mathbf{c} + \lambda\mathbf{d})^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0},$$

what is the maximum number of characteristic values of λ?

4. For the example of section 3.4 find the range of values of each of b_1, b_2, b_3 in turn for which the basic variables at optimality are unchanged.
5. For the example of section 3.4 find the range of values of each of c_2, c_3, c_4 in turn for which the optimum solution is unchanged.

NOTES

CHAPTER 9

THE SHOR-KHACHIAN ELLIPSOID METHOD

9.1

An important development in linear programming theory is the method due to N. Z. Shor and L. G. Khachian which leads to a polynomial-time algorithm in contrast with the simplex method which yields an exponential-time algorithm.

This comparison is discussed in this section, while some details of the method itself are described, and some of its properties established, in section 9.4. In section 9.2 we see how the ellipsoid method, which is directly concerned with finding a solution of a system of strict inequalities $\mathbf{A}\mathbf{x} < \mathbf{b}$, can be used to solve *l.p.p.s.* The method itself, despite its great theoretical interest, is unlikely to reduce the dominance of the simplex method as the approach for solving *l.p.p.s* in practice, so we do not discuss its practical implementation. Instead, as the method involves constructing a sequence of ellipsoids in n-space, this and other aspects of the background linear algebra are discussed briefly in section 9.3.

As we saw in section 2.9, for a *l.p.p.* with n variables, there could be as many as $n!/(m!(n-m)!)$ stages in the simplex method. As n increases, $n!$ increases like $(2\pi n)^{1/2}(n/e)^n$ (this is Stirling's approximation to $n!$). In a worst possible case, the simplex method could take as many stages as this, and so a definite upper bound on the time, or amount of work, required will involve the factor $(n/e)^n$.

When the amount of time possibly required by an algorithm involves the number of variables as an exponent it is said to be an exponential-time algorithm. In contrast, the amount of time required for the method of Gaussian elimination, for example, for solving a system of n linear equations in n variables increases with n like n^3 (see Appendix 3). This is an example of a polynomial-time algorithm, where n appears in the expression for the time required with a fixed exponent independent of n. Both the expressions $(n/e)^n$ and n^3 increase rapidly with n, but $(n/e)^n$ increases very much more rapidly; with $n = 100$ their

respective values are approximately 10^{157} and 10^6, and with $n = 1000$, 10^{2566} and 10^9.

The usual way to compare the amounts of time or amounts of work needed by different algorithms to solve the same problem is to evaluate $T(n)$, the number of arithmetic operations (additions and multiplications) each needs (see Appendix 3 for an example). $T(n)$ increases so rapidly with n for exponential-time algorithms that they soon become impractical for even the fastest computers. This means that the development of a polynomial-time algorithm for *l.p.p.s.* was a major mathematical goal. However these considerations are somewhat theoretical and in practice two other aspects are highly relevant.

The first is that, although the *nature* of $T(n)$ is of greatest importance, the constant or other factors multiplying the dominant term are also important. For the ellipsoid method the bound $T(n) = 4(n+1)^2 L \times \alpha n(m+n+\beta)$ can be established, where $\alpha n(m+n+\beta)$ is a bound for the number of operations required at each stage and α is small (see {15}). (The number of operations required by each *stage* of the simplex method (see section 7.2) is essentially $m(n-m)$.) The number L appears frequently in the analysis of the ellipsoid method. It is approximately the total number of binary digits in all the non-zero coefficients involved in the *l.p.p.* and can clearly be very large.

The second aspect is the way in which the amount of time suggested by $T(n)$ compares with the amount of time actually taken in practice. In cases where $T(n)$ is almost always very pessimistic its practical relevance may be slight. This is so for the simplex method, where the number of stages rarely exceeds a small multiple of m and can be expected to be nothing like exponential in n. For the ellipsoid method also, the bound given above may be somewhat pessimistic in practice, but the number L is explicitly involved in the algorithm (see section 9.4); and because the simplex method performs so efficiently (for general *l.p.p.s* at least) the ellipsoid method is not a practical alternative and is unlikely to have the impact one might at first expect of a polynomial-time algorithm.

9.2

The Shor-Khachian method finds a solution of a system of strict linear inequalities $\mathbf{Ax} < \mathbf{b}$ (if solutions exist), for the case where the elements of $\mathbf{A}$ and $\mathbf{b}$ are integers. The restriction to integer coefficients is crucial for establishing finiteness of the algorithm (and

hence its polynomial-time property) but, as we shall see in section 9.4, integer coefficients are not necessary for the operations of the algorithm itself. In practice any *l.p.p.* solved on a computer may be regarded as having integer coefficients because all the coefficients when stored must have a finite number of binary digits, and so multiplication of the (stored) constraints by appropriate powers of 2 would convert the coefficients to integers but leave the feasible region unchanged. Of course if the coefficients involved do not have an appropriate finite binary representation then the rounding-off that is required is equivalent to a perturbation of the problem; but we have seen in chapter 8 (and see also Appendix 3) that the effect of such a perturbation could be examined if necessary.

To write a *l.p.p.* as a single system of inequalities, without an objective function, we make use of the duality theorem. Suppose the problem is in standard primal form

$$\textit{minimise } \mathbf{c}^T\mathbf{x} \textit{ subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq \mathbf{0}. \tag{1}$$

This problem has a solution if and only if the dual problem

$$\textit{maximise } \mathbf{y}^T\mathbf{b} \textit{ subject to } \mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T,\ \mathbf{y} \geq \mathbf{0} \tag{2}$$

has a solution.

Thus the *l.p.p.* (1) has a solution if and only if the combined inequalities

$$(\mathbf{A}, \mathbf{O})\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \geq \mathbf{b},\ (\mathbf{O}, \mathbf{A}^T)\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \leq \mathbf{c},\ \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \geq \mathbf{0} \tag{3}$$

have a solution for $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}$. A solution $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}$ of (3) does not necessarily provide the optimum solutions of (1) and (2) unless we involve the condition $\mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{b}$, which is satisfied by optimum solutions. Since $\mathbf{c}^T\mathbf{x} - \mathbf{y}^T\boldsymbol{b} \geq 0$ for any feasible solutions $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}$ of (3), the requirement $\mathbf{c}^T\mathbf{x} - \mathbf{y}^T\mathbf{b} \leq 0$ restricts us to optimum solutions of (1) and (2).

Thus the constraints

$$(\mathbf{A}, \mathbf{O})\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \geq \mathbf{b},\ (\mathbf{O}, \mathbf{A}^T)\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \leq \mathbf{c},\ \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \geq \mathbf{0},\ \mathbf{c}^T\mathbf{x} - \mathbf{y}^T\mathbf{b} \leq 0 \tag{4}$$

have a solution $\begin{pmatrix}\mathbf{x}_0\\ \mathbf{y}_0\end{pmatrix}$ if and only if the *l.p.p.* (1) has a solution, and $\mathbf{x}_0$ and $\mathbf{y}_0$ are optimum solutions of (1) and (2) respectively.

Written as a single system of inequality constraints (4) becomes

$$\begin{pmatrix}-\mathbf{A} & \mathbf{O}\\ \mathbf{O} & \mathbf{A}^T\\ \mathbf{c}^T & -\mathbf{b}^T\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \leq \begin{pmatrix}-\mathbf{b}\\ \mathbf{c}\\ 0\end{pmatrix},\ \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \geq \mathbf{0}$$

or

$$\begin{pmatrix} -\mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}^T \\ \mathbf{c}^T & -\mathbf{b}^T \\ -\mathbf{I}_n & \mathbf{O} \\ \mathbf{O} & -\mathbf{I}_m \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \le \begin{pmatrix} -\mathbf{b} \\ \mathbf{c} \\ 0 \\ \mathbf{0}_n \\ \mathbf{0}_m \end{pmatrix}, \tag{5}$$

i.e. $\tilde{\mathbf{A}}\tilde{\mathbf{x}} \le \tilde{\mathbf{b}}$ where $\tilde{\mathbf{A}}$ is $(m+n+1+n+m)\times(n+m)$

In general the set of solutions of (5) will be a single point in $(m+n)$-space, whereas the set of solutions R of a feasible system of strict inequalities

$$\mathbf{A}'\mathbf{x} < \mathbf{b}'$$

is an open set with infinitely many points and a non-zero volume $V(R)$. (If $\mathbf{A}'\mathbf{x}_0 < \mathbf{b}'$, then $\mathbf{A}'(\mathbf{x}_0 + \delta\mathbf{x}) < \mathbf{b}'$ for *any* $\delta\mathbf{x}$ sufficiently small.)

That the magnitude of $V(R)$ is strictly positive is another aspect that is crucial for establishing the polynomial-time property of the ellipsoid method, as we explain in section 9.3, and so a given system $\tilde{\mathbf{A}}\tilde{\mathbf{x}} \le \tilde{\mathbf{b}}$ must be replaced by a system of strict inequalities. This can be done by perturbing $\tilde{\mathbf{b}}$ slightly, and it can be proved (see {15}) that

$$\tilde{\mathbf{A}}\tilde{\mathbf{x}} \le \tilde{\mathbf{b}},$$

where the elements of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ are integers, has a solution if and only if

$$\tilde{\mathbf{A}}\tilde{\mathbf{x}} < \tilde{\mathbf{b}} + 2^{-L}\mathbf{e}, \text{ where } \mathbf{e} = (1,1,\ldots,1)^T,$$

has a solution.

9.3

The Shor-Khachian method for finding a solution of

$$\mathbf{A}\mathbf{x} < \mathbf{b}, \tag{1}$$

where $\mathbf{A}$ is $m \times n$, is sequential, and at the k-th stage we have an n-vector $\mathbf{x}_k$ and an $n \times n$ matrix $\mathbf{B}_k$. These define an ellipsoid E_k in n-space with centre $\mathbf{x}_k$ which contains at least a part, S say, of the feasible region of (1). If $\mathbf{x}_k$ does not satisfy (1) then $\mathbf{x}_{k+1}$ and $\mathbf{B}_{k+1}$ are constructed so that the new ellipsoid E_{k+1} still contains the

whole of S but has volume $V(E_{k+1})$ which is less than $V(E_k)$, the volume of E_k. We will show that the volumes satisfy $V(E_{k+1})/V(E_k) \leq \gamma < 1$, where γ is a constant. Therefore $V(E_k)$ eventually becomes less than $V(S)$ if $\mathbf{x}_k$ never satisfies (1), which contradicts the existence of solutions of (1). We will also show that the ellipsoid E_k constructed at each stage does have the required properties. Some comments which provide background information about ellipsoids and affine transformations of n-space may be useful and these are given below. They require results from linear algebra that are not needed for the theory of the simplex method. These results are stated as they are needed, with brief comments but without proof; proofs and further explanation can be found in many texts on linear algebra, including {1}, {2} and {5}.

If E denotes the unit sphere with centre the origin,

$$E = \{\mathbf{x} \mid \mathbf{x}^T\mathbf{x} \leq 1\},$$

then for any non-singular matrix $\mathbf{Q}$

$$E_Q = \{\mathbf{Qx} \mid \mathbf{x} \in E\} = \{\mathbf{Qx} \mid \mathbf{x}^T\mathbf{x} \leq 1\}$$

is an ellipsoid in n-space with centre the origin. Alternatively $E_Q = \{\mathbf{x} \mid \mathbf{x}^T\mathbf{Q}^{-1T}\mathbf{Q}^{-1}\mathbf{x} \leq 1\} = \{\mathbf{x} \mid \mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} \leq 1\}$, where $\mathbf{Q}^{-1T}\mathbf{Q}^{-1} = \mathbf{B}^{-1}$, i.e. $\mathbf{B} = \mathbf{QQ}^T$, because if $\mathbf{x} \in E$, then $\mathbf{x}^T\mathbf{x} \leq 1$ and considering the vector $\mathbf{Qx}$,

$$(\mathbf{Qx})^T\mathbf{B}^{-1}(\mathbf{Qx}) = \mathbf{x}^T\mathbf{Q}^T\mathbf{Q}^{-1T}\mathbf{Q}^{-1}\mathbf{Qx} = \mathbf{x}^T\mathbf{x} \leq 1.$$

Any matrix $\mathbf{B}$ of the form $\mathbf{Q}^T\mathbf{Q}$, where $\mathbf{Q}$ is non-singular, is symmetric and positive definite (*ER*). It represents the ellipsoid E_Q which is the transformation T of the unit sphere E, where $T(\mathbf{x}) = \mathbf{Qx}$. Similarly, any symmetric positive-definite matrix $\mathbf{B}$ represents an ellipsoid $\{\mathbf{x} \mid \mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} \leq 1\}$ because there exists a non-singular matrix $\mathbf{Q}$ such that $\mathbf{QQ}^T = \mathbf{B}$. The matrix $\mathbf{Q}$ can be expressed in terms of the eigenvalues λ_j and eigenvectors $\mathbf{y}_j$ of $\mathbf{B}$, $j = 1,2, \ldots, n$. Because $\mathbf{B}$ is symmetric and positive-definite all its eigenvalues satisfy $\lambda_j > 0$ and it has a corresponding set of independent mutually orthogonal eigenvectors $\mathbf{y}_j$, $j = 1,2, \ldots, n$, satisfying

$$\mathbf{By}_j = \lambda_j\mathbf{y}_j,\ \mathbf{y}_i^T\mathbf{y}_j = 0 \text{ if } i \neq j \text{ and } \mathbf{y}_j^T\mathbf{y}_j = 1, \text{ for } i,j = 1,2, \ldots, n.$$

If $\mathbf{y}_j$ is the j-th column of the $n \times n$ matrix $\mathbf{Y}$, then $\mathbf{Y}^T = \mathbf{Y}^{-1}$, and with $\mathbf{D}$ the diagonal matrix with $d_{jj} = \lambda_j$, we have $\mathbf{BY} = \mathbf{YD}$;

so $$\mathbf{B} = \mathbf{YDY}^T = \mathbf{YD}^{1/2}\mathbf{D}^{1/2}\mathbf{Y}^T = \mathbf{QQ}^T,$$

where $\mathbf{Q} = \mathbf{YD}^{1/2}$ and $\mathbf{D}^{1/2}$ is the diagonal matrix with $(\mathbf{D}^{1/2})_{jj} = \lambda_j^{1/2}$, for $j = 1,2, \ldots, n$.

Notice that $\mathbf{B}^{-1}$ is also symmetric and positive definite, so it represents an ellipsoid

$$E_{Q^{-1}} = \{\mathbf{x} | \mathbf{x}^T\mathbf{B}\mathbf{x} \leq 1\},$$

with $\mathbf{B}^{-1} = (\mathbf{Q}^T)^{-1}\mathbf{Q}^{-1} = (\mathbf{Y}\mathbf{D}^{-1/2})(\mathbf{D}^{-1/2}\mathbf{Y}^T) = \mathbf{Y}\mathbf{D}^{-1}\mathbf{Y}^T$.

If we now consider the n mutually orthogonal unit vectors $\mathbf{e}_j$, $j = 1,2, \ldots, n$, which are the axes of the (spherical) ellipsoid E, then

$$\mathbf{Q}\mathbf{e}_j = \mathbf{Y}\mathbf{D}^{1/2}\mathbf{e}_j = \mathbf{Y}(\lambda_j^{1/2}\,\mathbf{e}_j) = \lambda_j^{1/2}\mathbf{Y}\mathbf{e}_j = \lambda_j^{1/2}\mathbf{y}_j.$$

So the effect of the transformation T, $T(\mathbf{x}) = \mathbf{Q}\mathbf{x}$, is to transform the axes of E into the mutually orthogonal unit eigenvectors $\mathbf{y}_j$ of $\mathbf{B}$, which are the directions of the axes of E_Q, and to "stretch" them by the factors $\lambda_j^{1/2}$. Thus the volume of E_Q, $V(E_Q)$, will be $\lambda_1^{1/2}\,\lambda_2^{1/2} \ldots \lambda_n^{1/2}\, V(E)$. Now

$$\lambda_1\lambda_2 \ldots \lambda_n = det(\mathbf{B}) = det(\mathbf{Q}\mathbf{Q}^T) = det(\mathbf{Q})det(\mathbf{Q}^T) = (det(\mathbf{Q}))^2,$$

so that $$V(E_Q) = |det(\mathbf{Q})|\, V(E).$$

If a translation by a vector $\mathbf{z}$ is added to the linear transformation $T(\mathbf{x}) = \mathbf{Q}\mathbf{x}$ we have an invertible affine transformation T, $T(\mathbf{x}) = \mathbf{Q}\mathbf{x} + \mathbf{z}$ (invertible because $\mathbf{Q}^{-1}$ exists and so the inverse transformation exists (see (2) below)). This transformation T maps E onto the ellipsoid $E_{Q,z}$ with centre $\mathbf{z}$,

$$E_{Q,z} = \{\mathbf{x} | (\mathbf{x} - \mathbf{z})^T\mathbf{B}^{-1}\,(\mathbf{x} - \mathbf{z}) \leq 1\} \quad \text{where } \mathbf{B} = \mathbf{Q}\mathbf{Q}^T,$$

because if $\mathbf{x}_0 \in E$, $\mathbf{x}_0^T\,\mathbf{x}_0 \leq 1$ and then

$$((\mathbf{Q}\mathbf{x}_0 + \mathbf{z}) - \mathbf{z})^T\mathbf{B}^{-1}\,((\mathbf{Q}\mathbf{x}_0 + \mathbf{z}) - \mathbf{z}) \leq 1$$

and so $T(\mathbf{x}_0) \in E_{Q,z}$. It is convenient to write $T(E)$ for $E_{Q,z}$. The translation of the ellipsoid E_Q onto $E_{Q,z}$ does not affect its volume so we have $V(E_{Q,z}) = det(\mathbf{Q})\, V(E)$.

The inverse T^{-1} of the transformation T is defined by

$$T^{-1}\,(\mathbf{x}) = \mathbf{Q}^{-1}\,(\mathbf{x} - \mathbf{z}) = \mathbf{Q}^{-1}\mathbf{x} - \mathbf{Q}^{-1}\mathbf{z} \tag{2}$$

because $$T(T^{-1}(\mathbf{x})) = T^{-1}\,(T(\mathbf{x})) = \mathbf{x} \quad (ER).$$

Notice that if $$E = \{\mathbf{x} | \mathbf{x}^T\mathbf{x} \leq 1\},$$

then $$T(E) = \{\mathbf{x} | (T^{-1}(\mathbf{x}))^T\, T^{-1}\,(\mathbf{x}) \leq 1\};$$

and generally, if $$E_o = \{\mathbf{x} | (\mathbf{x} - \mathbf{z}_0)^T\, \mathbf{B}_0^{-1}\,(\mathbf{x} - \mathbf{z}_0) \leq 1\}, \tag{3}$$

then $$T(E_o) = \{\mathbf{x} | (T^{-1}(\mathbf{x}) - \mathbf{z}_0)^T\, \mathbf{B}_0^{-1}(T^{-1}\,(\mathbf{x}) - \mathbf{z}_0) \leq 1\} \tag{4}$$

$$= \{\mathbf{x} | (\mathbf{x} - \mathbf{z} - \mathbf{Q}\mathbf{z}_0)^T\mathbf{Q}^{-1^T}\,\mathbf{B}_0^{-1}\,\mathbf{Q}^{-1}\,(\mathbf{x} - \mathbf{z} - \mathbf{Q}\mathbf{z}_0) \leq 1\}, \tag{5}$$

so $\mathbf{B}_0$ represents the ellipsoid E_o which has centre $\mathbf{z}_0$, and $\mathbf{Q}\mathbf{B}_0\mathbf{Q}^T$ represents the ellipsoid $T(E_o)$ which has centre $\mathbf{z} + \mathbf{Q}\mathbf{z}_0$.

Also $\mathbf{Q}\mathbf{B}_0\mathbf{Q}^T$ is symmetric and positive-definite if $\mathbf{B}_0$ is symmetric and positive-definite, and if $S \in E_o$ then $T(S) \in T(E_o)$ (ER).

9.4 The Shor-Khachian Algorithm

At the k-th stage the n-vector $\mathbf{x}_k$ and the $n \times n$ matrix $\mathbf{B}_k$ define an ellipsoid E_k with centre $\mathbf{x}_k$. If $\mathbf{x}_k$ does not satisfy $\mathbf{Ax} < \mathbf{b}$ then for some i, $1 \le i \le m$, $\mathbf{a}_{i*}^T \mathbf{x} \ge b_i$; and, writing $\mathbf{a}$ for $\mathbf{a}_{i*}$, we replace $\mathbf{x}_k$ and $\mathbf{B}_k$ by $\mathbf{x}_{k+1}$ and $\mathbf{B}_{k+1}$ where

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{n+1} \frac{\mathbf{B}_k \mathbf{a}}{(\mathbf{a}^T \mathbf{B}_k \mathbf{a})^{1/2}}, \tag{1}$$

$$\mathbf{B}_{k+1} = \frac{n^2}{n^2 - 1} \left(\mathbf{B}_k - \frac{2}{n+1} \frac{(\mathbf{B}_k \mathbf{a})(\mathbf{B}_k \mathbf{a})^T}{(\mathbf{a}^T \mathbf{B}_k \mathbf{a})} \right). \tag{2}$$

The expressions (1) and (2) are a convenient definition of the algorithm but are not the most suitable for practical implementation. Other versions slightly improve the efficiency and improve the numerical stability (see {16}). The expression $(\mathbf{B}_k\mathbf{a})\,(\mathbf{B}_k\mathbf{a})^T$ is an $n \times n$ matrix with rank 1 so that, apart from the factor $n^2(n^2-1)^{-1}$, $\mathbf{B}_{k+1}$ is obtained from $\mathbf{B}_k$ by a rank-one modification (this is a frequent device in non-linear optimisation algorithms).

Now consider the hyperplane $\{\mathbf{x} | \mathbf{a}^T(\mathbf{x} - \mathbf{x}_k) = 0\}$. This contains the centre of E_k and therefore separates E_k into two halves

$$E_{k-} \quad \text{in which} \quad \mathbf{a}^T(\mathbf{x} - \mathbf{x}_k) \le 0, \quad \text{and}$$
$$E_{k+} \quad \text{in which} \quad \mathbf{a}^T(\mathbf{x} - \mathbf{x}_k) > 0.$$

If $\mathbf{a}^T\mathbf{x}_k > b_i$ then $\mathbf{a}^T\mathbf{x} > b_i$ in the whole of E_{k+} so that E_{k-} contains the whole of the set S contained in E_k. The formulae (1) and (2) ensure that E_{k+1} contains the whole of E_{k-}. The validity of the construction (1) and (2) is established in theorem 11: the geometrical decrease in volume at each stage and hence convergence of the algorithm is established in theorem 12.

The algorithm begins with E_0 defined by $\mathbf{x}_0 = \mathbf{0}$ and $\mathbf{B}_0 = 2^{2L}\mathbf{I}$. It can be shown (see {15}) that if $\mathbf{Ax} < \mathbf{b}$ has any solutions then the set of solutions S contained in E_0 has volume at least $2^{-(n+1)L}$.

It is extremely helpful, before proving theorems 11 and 12, to simplify the situation by replacing the general ellipsoid E_k, represented by $\mathbf{x}_k$ and $\mathbf{B}_k$, and subsequent ellipsoid E_{k+1}, represented by $\mathbf{x}_{k+1}$ and $\mathbf{B}_{k+1}$, by E' and E^*, represented by $\mathbf{x}'$ and $\mathbf{B}'$, and $\mathbf{x}^*$ and $\mathbf{B}^*$ respectively, where E', $\mathbf{x}'$ and $\mathbf{B}'$ have a particularly simple form, namely $\mathbf{x}' = \mathbf{0}$ and $\mathbf{B}' = \mathbf{I}$, so that E' is just the unit sphere with centre the origin. To achieve this we apply an invertible affine transformation T which maps E_k onto E', so that

$$T(E_k) = E', \text{ and we write } T(E_{k+1}) = E^*.$$

We can choose T so that the vector $\mathbf{a}_{i*}$, involved at the k-th stage, becomes $-\mathbf{e}_1$ and then instead of $\mathbf{a}_{i*}$, $\mathbf{x}_k$, $\mathbf{B}_k$, $\mathbf{x}_{k+1}$, and $\mathbf{B}_{k+1}$ the properties of the Shor-Khachian algorithm can be established using $-\mathbf{e}_1$, $\mathbf{0}$, $\mathbf{I}$, $\mathbf{x}^*$, and $\mathbf{B}^*$, where $\mathbf{x}^* = T(\mathbf{x}_{k+1})$ and $\mathbf{B}^*$ is the matrix which defines E^*. The use of the transformation T is a valid device because the three properties we are concerned with,

$\mathbf{B}_{k+1}$ symmetric and positive-definite,
E_{k-} contained in E_{k+1}, and
$V(E_{k+1}) \leq \gamma\, V(E_k)$ where $\gamma < 1$,

are invariant under invertible affine transformations (see section 9.3 and exercise 9.1). We observe that T does not need to preserve the value of $\mathbf{a}^T\mathbf{x}_k$, because $\mathbf{a}$ merely defines a hyperplane perpendicular to $\mathbf{a}$ and containing the centre of E_k and hence the separation of E_k into E_{k-} and E_{k+}.

Substituting $\mathbf{0}$, $-\mathbf{e}_1$ and $\mathbf{I}$ in (1) and (2) for $\mathbf{x}_1^{(k)}$, $\mathbf{a}$ and $\mathbf{B}_k$ gives

$$\mathbf{x}^* = \frac{1}{n+1}\mathbf{e}_1, \tag{3}$$

$$\text{and} \quad \mathbf{B}^* = \frac{n^2}{n^2-1}\left(\mathbf{I} - \frac{2}{n+1}\mathbf{e}_1\mathbf{e}_1^T\right), \tag{4}$$

which is a diagonal matrix with diagonal elements

$$\left(\frac{1}{n+1}\right)^2, \frac{n^2}{n^2-1}, \frac{n^2}{n^2-1}, \ldots, \frac{n^2}{n^2-1}.$$

To verify that the transformation T which maps E_k onto E' also maps E_{k+1} onto the ellipsoid E^* defined by $\mathbf{x}^*$ and $\mathbf{B}^*$ of (3) and (4) we identify the inverse transformation T^{-1}. Writing $T^{-1} = \mathbf{Q}\mathbf{x} + \mathbf{z}$, $\mathbf{z}$ is clearly $\mathbf{x}_k$ and we see how $\mathbf{Q}$ can be found. Assuming $\mathbf{B}_k$ is symmetric and positive-definite (see theorem 11) then we know there is a non-singular matrix, $\mathbf{Q}_0$ say, such that $\mathbf{Q}_0\mathbf{Q}_0^T = \mathbf{B}_k$. Now let $\alpha = ((\mathbf{Q}_0^T\mathbf{a})^T(\mathbf{Q}_0^T\mathbf{a}))^{1/2}$ denote the length of $\mathbf{Q}_0^T\mathbf{a}$, so that $\mathbf{a}' = \alpha^{-1}\mathbf{Q}_0^T\mathbf{a}$ is a unit vector in the direction of $\mathbf{Q}_0^T\mathbf{a}$. An orthogonal matrix, $\mathbf{Q}_1$ say, can be found such that

$$\mathbf{Q}_1\mathbf{e}_1 = \mathbf{a}'. \tag{5}$$

The matrix $\mathbf{Q}_1$ represents the rotation of n-space which maps $-\mathbf{e}_1$ onto $\mathbf{a}'$, and we observe that

$$\alpha = ((\mathbf{Q}_1^T\mathbf{Q}_0^T\mathbf{a})^T(\mathbf{Q}_1^T\mathbf{Q}_0^T\mathbf{a}))^{1/2} = (\mathbf{a}^T\mathbf{B}_k\mathbf{a})^{1/2}$$

because $\mathbf{Q}_1\mathbf{Q}_1^T = \mathbf{I}$. The matrices $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are not unique and can be constructed in various ways. Their construction is not discussed

as T is being used purely as a convenient device for theoretical purposes and does not appear in the algorithm itself.

The required transformation T can now be defined by

$$T^{-1}(\mathbf{x}) = \mathbf{Q}_0\mathbf{Q}_1\,\mathbf{x} + \mathbf{x}_k$$

or

$$T(\mathbf{x}) = \mathbf{Q}_1^T\mathbf{Q}_0^{-1}(\mathbf{x} - \mathbf{x}_k) = \mathbf{Q}_1^T\mathbf{Q}_0^{-1}\,\mathbf{x} - \mathbf{Q}_1^T\mathbf{Q}_0^{-1}\,\mathbf{x}_k. \tag{6}$$

Using the formulae (3) and (4) we see that, with $T(\mathbf{x})$ given by (6) and $E_k = \{\mathbf{x}|(\mathbf{x} - \mathbf{x}_k)\mathbf{B}_k^{-1}(\mathbf{x} - \mathbf{x}_k)\}$, $T(E_k) = \{\mathbf{x}|\mathbf{x}^T\mathbf{x} \le 1\}$ (ER). Verifying that $T(E_{k+1}) = E^*$ is straightforward but worthwhile (see exercise 9.2).

We can now prove that the construction (1) and (2) for the Shor-Khachian algorithm is valid.

Theorem 11

(i) If $\mathbf{B}_k$ is symmetric and positive-definite then $\mathbf{B}_{k+1}$ given by (2) is symmetric and positive-definite.

(ii) The whole of E_{k-} is contained in E_{k+1} ■

Both these results can be verified in terms of E', E^*, $\mathbf{x}' = \mathbf{0}$, $\mathbf{B}' = \mathbf{I}$, and $\mathbf{x}^*$ and $\mathbf{B}^*$ given by (3) and (4).

From (4), $\mathbf{B}^*$ is diagonal and therefore symmetric; its diagonal elements are all positive so it is positive-definite.

As the vector which defines E'_- and E'_+ is $-\mathbf{e}_1$,

$$\begin{aligned} E' &= \{\mathbf{x}|\mathbf{x}^T\mathbf{x} \le 1, -\mathbf{e}_1^T\mathbf{x} \le 0\} \\ &= \{\mathbf{x}|\mathbf{x}^T\mathbf{x} \le 1, 0 \le \mathbf{x} \le 1\}. \end{aligned} \tag{7}$$

Let $\mathbf{x} \in E'_-$, then with $\mathbf{x}^* = \left(\dfrac{1}{n+1}\right)\mathbf{e}_1$ and

$$\mathbf{B}^{*-1} = dgl\left(\frac{(n+1)^2}{n^2}, \frac{n^2-1}{n^2}, \frac{n^2-1}{n^2}, \ldots, \frac{n^2-1}{n^2}\right),$$

$$\begin{aligned} (\mathbf{x} - \mathbf{x}^*)^T\mathbf{B}^{*-1}(\mathbf{x} - \mathbf{x}^*) &= \mathbf{x}^T\mathbf{B}^{*-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{B}^{*-1}\mathbf{x}^* + \mathbf{x}^{*T}\mathbf{B}^{*-1}\mathbf{x}^* \\ &= \frac{n^2-1}{n^2}\mathbf{x}^T\mathbf{x} + \left(\frac{(n+1)^2}{n^2} - \frac{n^2-1}{n^2}\right)x_1^2 \\ &\quad + \frac{(n+1)^2}{n^2}\left(\left(\frac{1}{n+1}\right)^2 - \frac{2x_1}{n+1}\right) \\ &= \frac{n^2-1}{n^2}(\mathbf{x}^T\mathbf{x} - 1) + \frac{n^2-1}{n^2} + \frac{1}{n^2} \\ &\quad + \frac{(n+1)}{n^2}2(x_1^2 - x_1) \end{aligned} \tag{8}$$

$$\leq 0 + 1 + \frac{2(n+1)}{n^2} x_1(x_1 - 1) \leq 1,$$

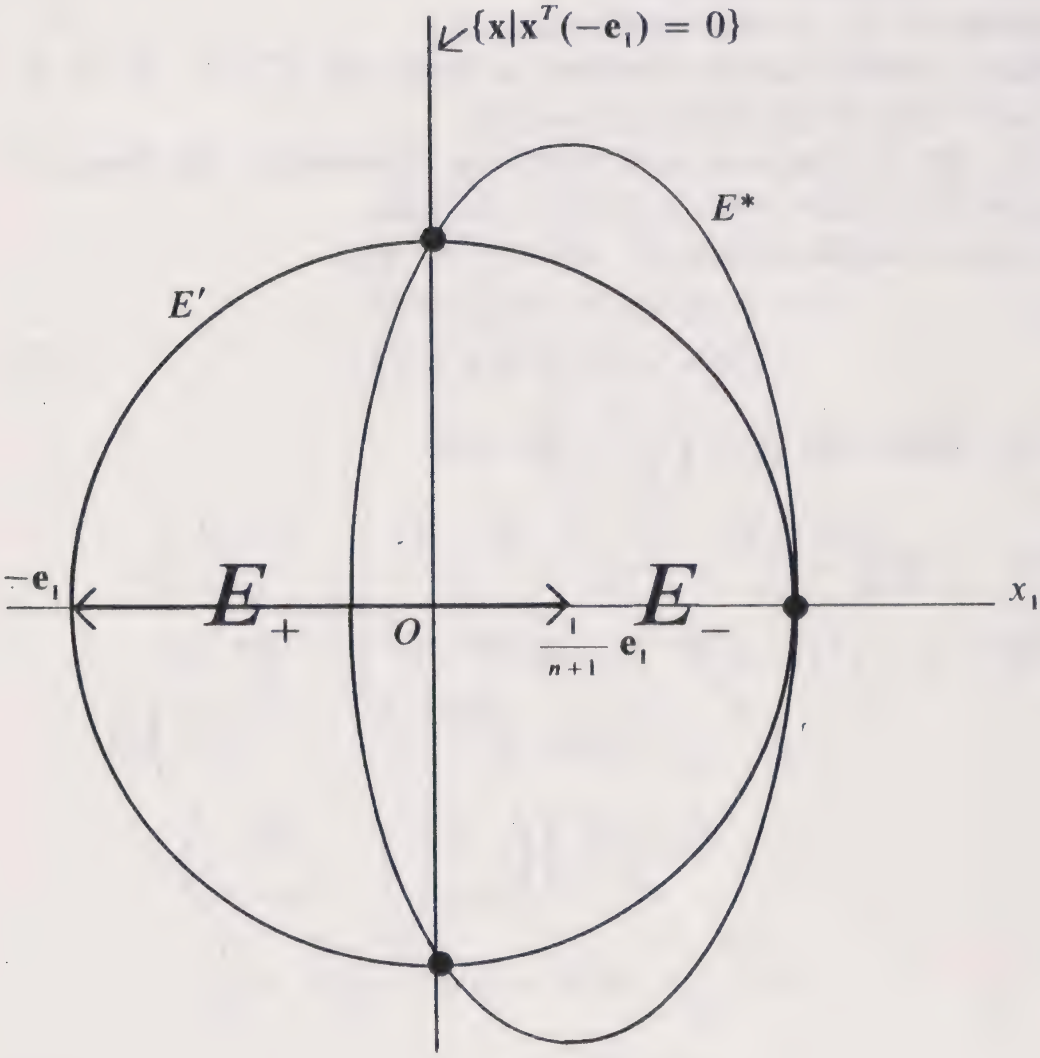
for $n \geq 2$ ■

The expression (8) enables us to describe the ellipsoid E^* geometrically, because $\mathbf{x}$ is on the boundary of E^* when

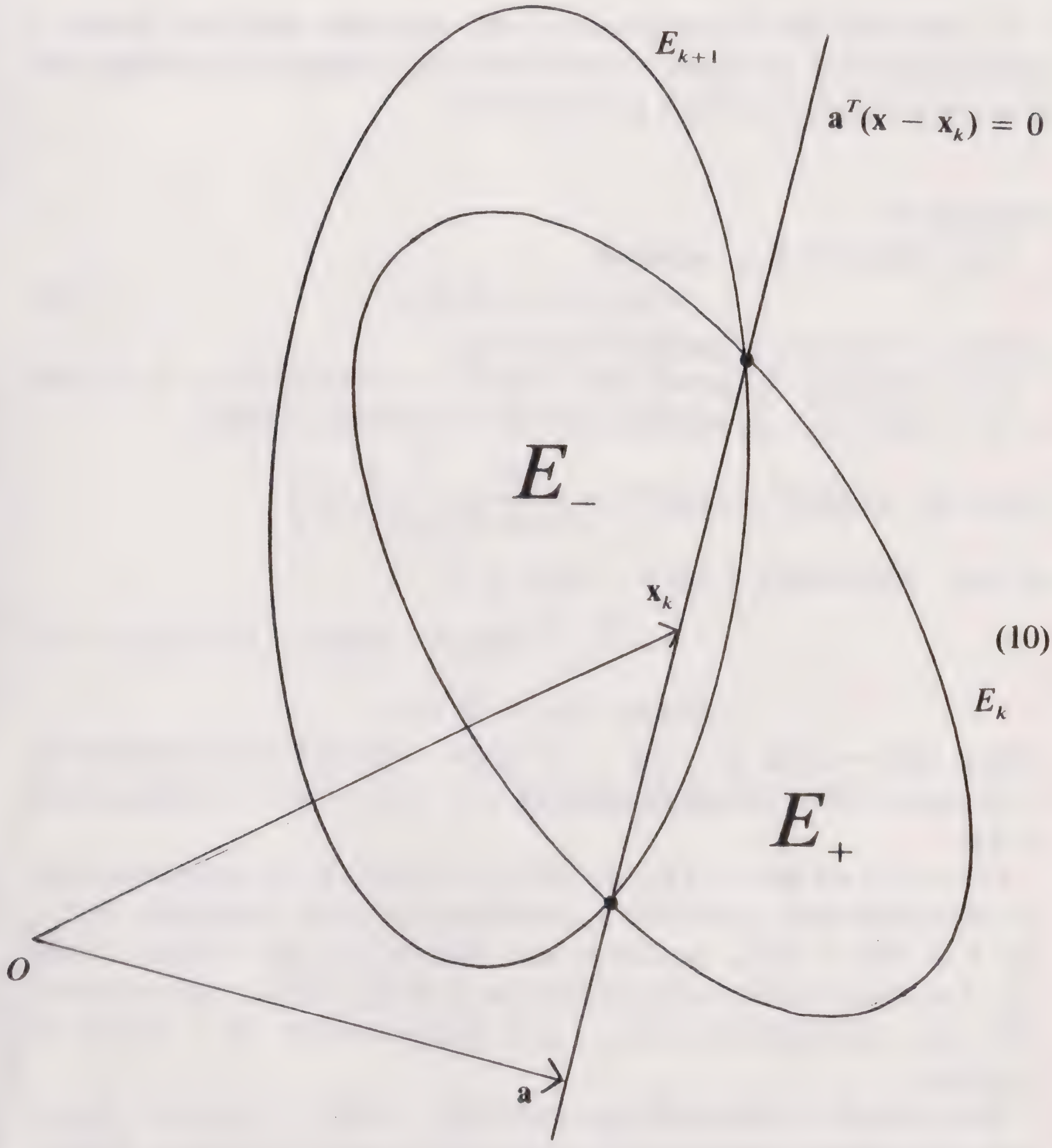
$$(\mathbf{x} - \mathbf{x}^*)^T \mathbf{B}^{*^{-1}} (\mathbf{x} - \mathbf{x}^*) = 1.$$

Therefore points $\mathbf{x}$ such that $\mathbf{x}^T\mathbf{x} = 1$ and $x_1 = 0$ are on the boundary of E^* and these points are also points in the intersection of E' and the hyperplane $\{\mathbf{x} | -\mathbf{e}_1^T\mathbf{x} = 0\}$ through the centre of E' perpendicular to $-\mathbf{e}_1$. A further isolated point, $x_1 = 1$, $\mathbf{x}^T\mathbf{x} = 1$, is common to both boundaries and so E' and E^* are tangential there.

In 2-space, or in a plane section through the origin and the x_1 axis, the situation is described by the diagram (9).



(9)



(10)

In general E_{k+1} intersects E_k in the $(n - 1)$-dimensional ellipsoid in which the hyperplane $\{\mathbf{x}|\mathbf{a}^T(\mathbf{x} - \mathbf{x}_k) = 0\}$ intersects E_k, and also intersects E_k, tangentially, at $T^{-1}(\mathbf{e}_1)$. The point

$$T^{-1}(\mathbf{e}_1) = \mathbf{Q}_0\mathbf{Q}_1\mathbf{e}_1 + \mathbf{x}_k = -\frac{1}{\alpha}\mathbf{B}_k\mathbf{a} + \mathbf{x}_k$$

is not (in general) in the plane of the diagram (10), which is the plane defined by $\mathbf{x}_k$ and $\mathbf{a}$.

To establish the convergence of the ellipsoid algorithm within a specified number of stages we prove that the volumes of the ellipsoids generated decrease at least geometrically.

Theorem 12

The volume of E_{k+1} satisfies

$$V(E_{k+1}) \leq \gamma\, V(E_k) \tag{11}$$

where $\gamma < 1$ and γ is independent of k■

It is sufficient to prove that $V(E^*) \leq \gamma V(E')$ for $\gamma < 1$, and as $\mathbf{B}^* = \mathbf{Q}\mathbf{Q}^T$ we can establish (11) by evaluating $|det(\mathbf{Q})|$.

From (4), $(det(\mathbf{Q}))^2 = det(\mathbf{B}^*) = \dfrac{n^2}{(n+1)^2}\left(\dfrac{n^2}{n^2-1}\right)^{n-1}$

so that
$$\begin{aligned} log|det(\mathbf{Q})| &= log\, n - log(n+1) \\ &\quad + \frac{n-1}{2}(log\, n^2 - log(n-1) - log(n+1)) \\ &= \tfrac{1}{2}(t(n) - t(n+1)) < 0, \end{aligned}$$

where $t(n) = n\, log\, n - (n-1)\, log(n-1)$ and $t(n)$ increases as n increases (ER). Therefore $|det(\mathbf{Q})| = \gamma < 1$, where γ is independent of k■

From (11) we have $V(E_k) \leq \gamma^k V(E_0)$, so that if the algorithm does not terminate with a feasible $\mathbf{x}_k$ satisfying $\mathbf{A}\mathbf{x}_k < \mathbf{b}$, eventually $V(E_k)$, for $k = 4(n+1)^2 L$, becomes less than $V(S)$, the volume of the set of solutions which are contained in E_0 if $\mathbf{A}\mathbf{x} < \mathbf{b}$ has any solutions. This is a contradiction and so if it happens then $\mathbf{A}\mathbf{x} < \mathbf{b}$ has no solutions.

The method of ellipsoids has drawbacks which, for general *l.p.p.s*, make the simplex method much superior for practical purposes. Firstly, if a *l.p.p.* is infeasible, to reveal this the prescribed, very large, number of stages is required. Secondly, and more importantly, the number γ is extremely close to 1, (more so as n increases) so that the ellipsoids generated may decrease in volume very slowly. In practice this is usually the case and the ellipsoids and their centroids that occur show no regular behaviour that could be used to predict the eventual outcome.

Exercises 9

1. Explain how a non-negative solution of a system of inequalities $\tilde{\mathbf{A}}\tilde{\mathbf{x}} \le \tilde{\mathbf{b}}$ can be found (see exercise 4.4).
2. Let E_k be the ellipsoid $\{\mathbf{x} | (\mathbf{x} - \mathbf{x}_k)^T \mathbf{B}_k^{-1} (\mathbf{x} - \mathbf{x}_k) \le 1\}$ and E_{k+} and E_{k-} the two halves of E_k defined by the hyperplane $H = \{\mathbf{x} | \mathbf{a}^T(\mathbf{x} - \mathbf{x}_k) = 0\}$. If T is the affine transformation which maps E_k onto E', the unit sphere with centre the origin, verify that $T(H)$ defines two halves of E' which are $T(E_{k+})$ and $T(E_{k-})$.
3. Using the definition (6) in section 9.4 verify that the affine transformation T which maps E_k onto E' also maps E_{k+1} onto E^*.
4. For the case $n = 2$ obtain $\mathbf{x}^*$ and $\mathbf{B}^*$ and verify that diagram (9) describes this case. How does this diagram change as n increases?
5. Evaluate the amount of work (e.g. the number of multiplications) in one stage of the ellipsoid algorithm, taking into account the structure of the matrix involved (see (5) section 9.2), and compare it with the amount of work in one stage of the simplex method.
6. Does the choice of $\mathbf{a}_{i*}$ (see section 9.4) affect the number of stages needed in the ellipsoid method? Discuss the practical implications.
7. A "deeper cut" of the current ellipsoid E_k than that through the centroid $\mathbf{x}_k$ and defined by the hyperplane $\mathbf{a}_{i*}^T(\mathbf{x} - \mathbf{x}_k) = 0$ can be made so that E_{k-} is less than half of E_k. What is the best alternative hyperplane based on the violated constraint $\mathbf{a}_{*i}\mathbf{x} < \mathbf{b}_i$?

NOTES

CHAPTER 10

TRANSPORTATION AND SIMILAR PROBLEMS

10.1

The matrix of coefficients **A** of a transportation problem is very sparse and so such a problem would be a natural candidate for solution by the revised simplex method. However, as we saw in section 1.3 **A** is more than just sparse: there is a pronounced special structure of non-zero elements, all of which have value 1, and the structure is exactly the same for all transportation problems. This results in even more efficient algorithms in which the initial data is retained unchanged throughout. These algorithms are particularly interesting because although they can be defined without reference to the simplex method, they really consist of the simplex method performed implicitly, and also because the duality theorem and the dual problem play a crucial part.

The problem is to choose the amounts x_{ij} of some commodity to be transported from each of m sources $D_1, D_2, \ldots, D_m$ to each of n destinations $B_1, B_2, \ldots, B_n$ so that the total cost $\Sigma_{i=1}^{m} \Sigma_{j=1}^{n} c_{ij}\, x_{ij}$ is minimised.

For $i = 1, 2, \ldots, m$ the total amount taken from D_i, $\Sigma_{j=1}^{n} x_{ij}$, cannot exceed the amount d_i which is available there, and for $j = 1, 2, \ldots, n$ the total amount taken to B_j, $\Sigma_{i=1}^{m} x_{ij}$, should not be less than the amount b_j required there.

As we observed in section 1.3, if

$$\Sigma_{i=1}^{m} d_i = \Sigma_{j=1}^{n} b_j \tag{1}$$

then we have the *l.p.p.*

$$\begin{gathered} \textit{minimise} \quad \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} c_{ij}\, x_{ij} \quad \textit{subject to} \\ x_{ij} \geq 0, \ \Sigma_{j=1}^{n} x_{ij} = d_i, \ \Sigma_{i=1}^{m} x_{ij} = b_j, \\ i = 1, 2, \ldots, m, \ j = 1, 2, \ldots, n. \end{gathered} \tag{2}$$

In practice we are unlikely to have exact equality in (1) and so there will have to be a source with some surplus or a destination whose requirements are not met. In order to produce a *l.p.p.* in canonical form, if $\Sigma_{i=1}^{m} d_i < \Sigma_{j=1}^{n} b_j$ we introduce a fictitious source D_{m+1}

containing $d_{m+1} = \Sigma_{j=1}^{n} b_j - \Sigma_{i=1}^{m} d_i$ of the commodity, and if $\Sigma_{i=1}^{m} d_i > \Sigma_{j=1}^{n} b_j$ we introduce a fictitious destination B_{n+1} requiring $b_{n+1} = \Sigma_{j=1}^{m} d_i - \Sigma_{j=1}^{n} b_j$ of the commodity. In either case the corresponding fictitious transportation cost coefficients are all zero. We shall assume that this modification of the problem has already been made if necessary, so from now on we have $\Sigma_{i=1}^{m} d_i = \Sigma_{j=1}^{n} b_j$ and the *l.p.p.*

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{x} \geq \mathbf{0} \quad \textit{and} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \text{ where} \tag{3}$$

$$\mathbf{c}^T = (c_{11}, c_{12}, \dots c_{1n}, c_{21}, c_{22}, \dots, c_{2n} \dots, c_{m1}, c_{m2}, \dots, c_{mn}),$$

$$\mathbf{x}^T = (x_{11}, x_{12}, \dots, x_{1n}, x_{21}, x_{22}, \dots, x_{2n}, \dots, x_{m1}, x_{m2}, \dots x_{mn}),$$

$$\mathbf{b}^T = (d_1, d_2, \dots, d_m, b_1, b_2, \dots, b_n) \quad \text{and}$$

$\mathbf{A}$ is the $(m+n) \times (mn)$ matrix described in section 1.3.

Since the sum of the first m rows of $\mathbf{A}$ and the sum of the last n rows of $\mathbf{A}$ are the same, $\mathbf{A}$ has rank less than $(m+n)$ (*ER*), and in fact $r(\mathbf{A}) = (m+n-1)$ (see exercise 10.4).

The dual problem is

$$\textit{maximise} \quad \mathbf{y}^T\mathbf{b} \quad \textit{subject to} \quad \mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T, \tag{4}$$

where $\mathbf{y}$ is an $(m+n)$-vector. It is helpful to write $\mathbf{y}$ as

$$\mathbf{y} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix},$$

where $\mathbf{u}$ is an m-vector and $\mathbf{v}$ is an n-vector.

Then (4) can be written

$$\textit{maximise} \quad \Sigma_{i=1}^{m} u_i d_i + \Sigma_{j=1}^{n} v_j b_j \quad \textit{subject to} \quad u_i + v_j \leq c_{ij},$$
$$i = 1, 2, \dots, m, \; j = 1, 2, \dots, n. \tag{5}$$

As there are only $(m+n-1)$ independent primal equality constraints, and omitting any one of the $(m+n)$ equality constraints gives an independent set, we really only need $(m+n-1)$ dual variables. We could for example omit the first primal equality constraint and omit u_1. Instead we retain both, to preserve the symmetry of the problems, but we have only $(m+n-1)$ basic variables in a *b.f.s.* of the primal, and consequently we always set one dual variable to zero.

10.2

A method for solving transportation problems is developed by solving, in an intuitive fashion, a particular example. Specific parts of the method that emerges are discussed in more detail in the following sections.

Let $m = 3$, $n = 4$; $d_1 = 4$, $d_2 = 4$, $d_3 = 8$; $b_1 = 3$, $b_2 = 6$, $b_3 = 4$, $b_4 = 4$, and the cost coefficients c_{ij} be given by the cost matrix $\mathbf{C}$,

$$\mathbf{C} = \begin{pmatrix} 2 & 3 & 4 & 3 \\ 4 & 3 & 2 & 1 \\ 2 & 4 & 5 & 4 \end{pmatrix}. \tag{1}$$

In this example $\Sigma_i d_i < \Sigma_j b_j$, so we introduce a fourth source D_4 containing 1 unit of the commodity, $d_4 = 1$, $c_{4j} = 0$, $j = 1, 2, 3, 4$, and from now on $m = 4$, $n = 4$.

The values x_{ij} for any chosen solution $\mathbf{x}$ themselves constitute an $m \times n$ matrix $\mathbf{X}$. The sum of the rows of $\mathbf{X}$ must be $\mathbf{b}$ and the sum of the columns must be $\mathbf{d}$. This leads to a simple method for finding an initial *b.f.s.*, called the *northwest corner method.*

3	1			4
	4			4
	1	4	3	8
			1	1
3	6	4	4	b_j \ d_i

(2)

Starting with x_{11}, the northwest corner element of $\mathbf{X}$, we put $x_{11} = min\ (b_1, d_1)$, which is $b_1 = 3$ in this case. This means that all other elements of the first column of $\mathbf{X}$ must be zero, so this column can be *removed,* and the *remaining row sum* of the first row of the remaining part of $\mathbf{X}$ is $4 - 3 = 1$. This principle is now repeated with the remaining parts of $\mathbf{X}$, $\mathbf{b}$ and $\mathbf{d}$. Thus $x_{12} = min\ (6, 1) = 1$, $x_{22} = min\ (5, 4) = 4$ and so on. Each step determines the remaining elements of one row or one column of $\mathbf{X}$, except the very last choice x_{mn} which completes the m-th row and the n-th column of $\mathbf{X}$, so that in general $(m + n - 1)$ elements of $\mathbf{X}$, i.e., of $\mathbf{x}$, will be assigned a non-zero value.

The same method for the case $\mathbf{b} = (4, 5, 4, 4)$ and $\mathbf{d} = (4, 4, 8, 1)$ gives

4				4
	4			4
	1	4	3	8
			1	1
4	5	4	4	b_j \ d_i

(3)

and one of the zero x_{ij} must be chosen as a basic variable. We will adopt the convention that, in this situation, x_{12} or x_{21} is chosen; it does not matter which, and the one with the smaller cost coefficient would be a natural choice.

The value of this solution (2) (i.e. the cost of this particular transportation scheme) is

$$\mathbf{c}^T\mathbf{x} = 2\times 3+3\times 1+3\times 4+4\times 1+5\times 4+4\times 3 = 57.$$

By the equilibrium theorem for canonical form, we know that if the *b.f.s.* we have just obtained is optimum, then the dual constraints corresponding to basic variables are satisfied as equalities (see exercises 5.8 and 6.8). Using this result to determine the vectors **u** and **v** gives the seven equations:

$$u_1+v_1=2,\ u_1+v_2=3,\ u_2+v_2=3,\ u_3+v_2=4,$$
$$u_3+v_3=5,\ u_3+v_4=4,\ u_4+v_4=0. \tag{4}$$

Imposing the additional equation $u_1=0$ determines **u** and **v**; $v_1=2$, $v_2=3$, $u_2=0$, $u_3=1$, $v_3=4$, $v_4=3$, $u_4=-3$, and this computation can conveniently be performed in a compact tableau similar to and using (2).

v_j \ u_i	2	3	4	3
0	2	3		
0		3		
1		4	5	4
−3				0

(5)

Suppose we now evaluated u_i+v_j for x_{ij} non-basic and found that $u_i+v_j \le c_{ij}$, $i=1,2,\dots,m$, $j=1,2,\dots,n$. Then **u** and **v** would satisfy

the dual constraints (4) or (5) of section 10.1, and

$$(\mathbf{u}^T, \mathbf{v}^T)\begin{pmatrix}\mathbf{d}\\ \mathbf{b}\end{pmatrix} - \mathbf{c}^T\mathbf{x} = (\mathbf{u}^T, \mathbf{v}^T)\mathbf{A}\mathbf{x} - \mathbf{c}^T\mathbf{x}$$
$$= ((\mathbf{u}^T, \mathbf{v}^T)\mathbf{A} - \mathbf{c}^T)\mathbf{x}.$$

But for $x_{ij} \neq 0$, $((\mathbf{u}^T,\mathbf{v}^T)\mathbf{A} - \mathbf{c}^T)_{ij} = u_i + v_j - c_{ij} = 0$, so that $\mathbf{x}$ and $(\mathbf{u}^T, \mathbf{v}^T)$ are optimum solutions for the primal and dual *l.p.p.s* respectively, by the duality theorem.

If $\mathbf{u}$ and $\mathbf{v}$ do not satisfy all the dual constraints we attempt to find an improved *b.f.s.* of the primal by asking what would happen with the corresponding *b.f.s.* in the simplex method. The essential information we need is the vector of *e.c.c.s* $\mathbf{c}'$, which we obtain by adding to $\mathbf{c}$ multiples of rows of $\mathbf{A}$ so that c'_{ij} is zero if x_{ij} is basic. In the present situation, we do not have an equivalent system of primal constraints

$$\mathbf{A}'\mathbf{x} = \begin{pmatrix}\mathbf{d}'\\ \mathbf{b}'\end{pmatrix}$$

in which the columns of $\mathbf{A}'$ corresponding to basic variables x_{ij} are columns of the unit matrix, so the appropriate row multipliers are not just $-c_{ij}$. However the vector $\mathbf{c}'$ given by

$$\mathbf{c}'^T = \mathbf{c}^T - (\mathbf{u}^T, \mathbf{v}^T)\mathbf{A}$$

satisfies the conditions $c'_{ij} = 0$ if x_{ij} is basic.

So, for (i,j) such that x_{ij} is non-basic, we evaluate $((\mathbf{u}^T, \mathbf{v}^T)\mathbf{A})_{ij} = u_i + v_j$, and these values can be put in the empty cells of the tableau (5). Notice that these $u_i + v_j$ are just the w_{ij} of section 5.4 so it is natural to call them w_{ij}. The current situation can be described unambiguously by a single tableau if each cell contains

$\boxed{^{c_{ij}}\ x_{ij}}$ if x_{ij} is basic, or $\boxed{^{c_{ij}}\ {}_{w_{ij}}}$ if x_{ij} is non-basic.

$u_i \backslash v_j$	2	3	4	3	
0	$^{2}\ 3$	$^{3}\ 1$	$^{4}\ {}_{4}$	$^{3}\ {}_{3}$	4
0	$^{4}\ {}_{2}$	$^{3}\ 4$	$^{2}\ {}_{4}$	$^{1}\ {}_{3}$	4
1	$^{2}\ {}_{3}$	$^{4}\ 1$	$^{5}\ 4$	$^{4}\ 3$	8
−3	$^{0}\ {}_{-1}$	$^{0}\ {}_{0}$	$^{0}\ {}_{-1}$	$^{0}\ 1$	1
	3	6	4	4	$b_j \backslash d_i$

(6)

In the simplex method we introduce a positive amount θ of the non-basic variable x_r corresponding to the negative *e.c.c.* of largest magnitude, where θ is given the largest possible value consistent with maintaining feasibility while changing the other basic variable to still satisfy the equality constraints.

In the present situation the negative *e.c.c.s* are c'_{23}, c'_{24}, c'_{31} and c'_{43}. The largest in magnitude of these is c'_{23} (or c'_{24}), so we put $x_{23} = \theta$. To preserve the row and column sums we have successively to replace x_{22} by $x_{22} - \theta$, x_{32} by $x_{32} + \theta$, and x_{33} by $x_{33} - \theta$. From these replacement values we see that the maximum value for θ is 4, which leads to the following *b.f.s.*

3	1			4
		4		4
	5		3	8
			1	1
3	6	4	4	$b_j \setminus d_i$

(7)

As there are only six positive basic variables, we retain one of x_{22} and x_{33} as a basic variable with value zero, and since x_{22} has the smaller cost coefficient this is the natural choice.

The whole procedure can now be repeated with a single compact tableau as (6), which is constructed by inserting in order

b_i, d_j, c_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$,
x_{ij} for basic variables,
u_i, v_j, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$,
w_{ij} for x_{ij} non-basic,
$\checkmark$ where $c'_{ij} = c_{ij} - w_{ij} \geq 0$,
$\pm\theta$ starting in (s, t) cell where $\min\limits_{ij} c'_{ij} = c'_{st}$.

Before doing so, we check the value of the *b.f.s.* just obtained. This is $6 + 3 + 8 + 20 + 12 = 49 = 57 + \theta \times \min c'_{ij}$.

$u_i \backslash v_j$	2	3	2	3	
0	$^{2}\ 3$	$^{3}\ 1$	$^{4}\ \checkmark_{2}$	$^{3}\ \checkmark_{3}$	4
0	$^{4}\ \checkmark_{2}$	$^{3}\ 0-\theta$	$^{2}\ 4$	$^{1}\ \theta_{3}$	4
1	$^{2}\ \ _{3}$	$^{4}\ 5+\theta$	$^{5}\ \checkmark_{3}$	$^{4}\ 3-\theta$	8
-3	$^{0}\ \checkmark_{-1}$	$^{0}\ \checkmark_{0}$	$^{0}\ \checkmark_{-1}$	$^{0}\ 1$	1
	3	6	4	4	$b_j \backslash d_i$

Value $= 6 + 3 + 8 + 20 + 12 = 49$,
min $c'_{ij} = c'_{24} = -2$,
$\theta = 0$.
$c'_{24}\theta = 0$. (8)

$u_i \backslash v_j$	2	3	4	3	
0	$^{2}\ 3$	$^{3}\ 1$	$^{4}\ \checkmark_{2}$	$^{3}\ \checkmark_{3}$	4
-2	$^{4}\ \checkmark_{0}$	$^{3}\ \checkmark_{1}$	$^{2}\ 4-\theta$	$^{1}\ 0+\theta$	4
1	$^{2}\ \ _{3}$	$^{4}\ 5$	$^{5}\ \checkmark_{5}$	$^{4}\ 3$	8
-3	$^{0}\ \checkmark_{-1}$	$^{0}\ \checkmark_{0}$	$^{0}\ \theta_{1}$	$^{0}\ 1-\theta$	1
	3	6	4	4	$b_j \backslash d_i$

Value $= 6 + 3 + 8 + 20 + 12 = 49$,
min $c'_{ij} = c'_{31} = c'_{43} = -1$, choosing
c'_{43} because $c_{43} < c_{31}$ leads to
$\theta = 1$. (9)
$c'_{43}\theta = -1$.

$u_i \backslash v_j$	2	3	4	3	
0	$^{2}\ 3-\theta$	$^{3}\ 1+\theta$	$^{4}\ \checkmark_{4}$	$^{3}\ \checkmark_{3}$	4
-2	$^{4}\ \checkmark_{0}$	$^{3}\ \checkmark_{1}$	$^{2}\ 3$	$^{1}\ 1$	4
1	$^{2}\ \theta_{3}$	$^{4}\ 5-\theta$	$^{5}\ \checkmark_{5}$	$^{4}\ \checkmark_{3}$	8
-4	$^{0}\ \checkmark_{-2}$	$^{0}\ \checkmark_{-1}$	$^{0}\ 1$	$^{0}\ \checkmark_{-1}$	1
	3	6	4	4	$b_j \backslash d_i$

Value $= 6 + 3 + 6 + 1 + 20 + 12 = 48$,
min $c'_{ij} = c'_{31} = -1$,
$\theta = 3$.
$c'_{31}\theta = -3$. (10)

$u_i \backslash v_j$	2	3	4	3	
0	^{2} ✓ $_{1}$	^{3} 4	^{4} ✓ $_{4}$	^{3} ✓ $_{3}$	4
−2	^{4} ✓ $_{-1}$	^{3} ✓ $_{1}$	^{2} 3	^{1} 1	4
1	^{2} 3	^{4} 2	^{5} ✓ $_{5}$	^{4} 3	8
−4	^{0} ✓ $_{-3}$	^{0} ✓ $_{-1}$	^{0} 1	^{0} ✓ $_{-1}$	1
	3	6	4	4	$b_j \backslash d_i$

Value $= 12 + 6 + 1 + 6 + 8 + 12 = 45$,
all $c'_{ij} \geq 0$,
hence current *b.f.s.* is optimum. (11)

The dual variables $\mathbf{u}, \mathbf{v}$ satisfy the dual constraints $u_i + v_j \leq c_{ij}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and, as it should be, the value of the dual solution $\mathbf{u}^T\mathbf{d} + \mathbf{v}^T\mathbf{b}$ is

$$-8 + 8 - 4 + 3 + 18 + 16 + 12 = 45.$$

In this solution it is the third destination which receives less than its stated requirement.

10.3

Since the method which evolved in the previous section was just the simplex method we know it will reach the optimum solution in a finite number of steps. However, it is essential in the simplex method that the columns of **A** corresponding to the basic variables are independent, so we must check that this is the case initially and that the *θ-circuit* preserves this property. Three other pertinent aspects are discussed in the following section. These are:

(i) whether the northwest corner method provides a *good* initial *b.f.s.*,
(ii) whether $\mathbf{x}_{opt}$ necessarily has integer valued elements when **b** and **d** have integer valued elements, and
(iii) the question of cycling when *b.f.s.s* are degenerate.

To show that the columns of **A** corresponding to the basic variables determined by the northwest corner solution are independent we just emphasise the convention of section 10.2, that when the *remaining row sum* and the *remaining column sum* are the same, for example $b_1 = d_1$, or $b_1 > d_1$ and $b_1 - d_1 = d_2$, we will reduce the remaining part of **X** by removing its first row *or* column (it does not matter which, and in these instances the next basic variable chosen will

have value zero). Thus in every case the number of basic variables determined is $(m+n-1)$, there is (at least) one row or column of $\mathbf{X}$ with only one basic variable and, crucially, if we remove this row or column, then the remaining part of $\mathbf{X}$ has the same property (ER).

Denoting the corresponding columns of $\mathbf{A}$ by an $(m+n) \times (m+n-1)$ matrix $\mathbf{B}$, the northwest corner method has produced a solution of

$$\mathbf{B}\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{d} \\ \mathbf{b} \end{pmatrix}, \tag{1}$$

where $\tilde{\mathbf{x}}$ is the $(m+n-1)$-vector of basic variables. Since we have a solution, $\begin{pmatrix} \mathbf{d} \\ \mathbf{b} \end{pmatrix}$ is in the column space of $\mathbf{B}$ and so either $\mathbf{B}$ has full rank, i.e. $r(\mathbf{B}) = (m+n-1)$, or the solution $\tilde{\mathbf{x}}$ is not unique. But $\tilde{\mathbf{x}}$ *is* unique, because, returning to the equivalent situation in which row sums and column sums of $\mathbf{X}$ are equal to elements of $\mathbf{d}$ and $\mathbf{b}$, as there must be at least one row or column of $\mathbf{X}$ with only one basic x_{ij}, the value of this x_{ij} is uniquely determined. If we now remove this row or column of $\mathbf{X}$ exactly the same argument holds for the remaining $(m+n-1)$ rows and columns of $\mathbf{X}$ and $(m+n-2)$ basic variables. Hence, inductively, all the x_{ij} are uniquely determined, so that $\mathbf{B}\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{d} \\ \mathbf{b} \end{pmatrix}$ has a unique solution and $\mathbf{B}$ has full rank. We could regard the northwest corner method as identifying a set of $(m+n-1)$ basic variables, whose values $\mathbf{x}$ are then determined by (1). From the way in which the method chooses the basic variables, we can see that each can have only one value if they are to satisfy (1).

To show that successive *b.f.s.s do* correspond to independent columns of $\mathbf{A}$ we examine the procedure of the *θ-circuit*. This consists of alternate steps along rows and columns of $\mathbf{X}$ and must involve only rows and columns of $\mathbf{X}$ with at least two basic variables, the only possible exceptions being the first and last steps from the new variable, x_{st} say, that has just been selected. The complete circuit defines a closed path and identifies a number of the current basic variables together with x_{st}. The columns of $\mathbf{A}$ corresponding to the variables defining any such path are linearly dependent (see exercise 10.7). When the value of θ has been chosen, and x_{st} is given this value, one of the basic variables on the circuit, $x_{s't'}$ say, has value zero and all of them have a unique value. This follows because if $\mathbf{d}'$ denotes $\mathbf{d}$ with $d_{s'}$ replaced by $d_{s'} - \theta$, and $\mathbf{b}'$ denotes $\mathbf{b}$ with $b_{t'}$ replaced by $b_{t'} - \theta$, then $\mathbf{B}\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{d}' \\ \mathbf{b}' \end{pmatrix}$ still has a unique solution

(with $x_{s't'} = 0$). Thus if $\mathbf{B}'$ denotes $\mathbf{B}$ with the column corresponding to $x_{s't'}$ replaced by the column of $\mathbf{A}$ corresponding to x_{st}, and $\tilde{\mathbf{x}}'$ denotes $\tilde{\mathbf{x}}$ with $x_{s't'}$ replaced by x_{st}, then $\mathbf{B}'\tilde{\mathbf{x}}' = \begin{pmatrix} \mathbf{d} \\ \mathbf{b} \end{pmatrix}$ has a unique solution (with $x_{st} = \theta$) and so the columns of $\mathbf{B}'$ are independent. (Alternatively, see exercise 10.7.)

We observe at this point that the equations which determine $\mathbf{u}$ and $\mathbf{v}$ are

$$\mathbf{B}^T\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \mathbf{c}_1,$$

where $\mathbf{c}_1$ is an $(m + n - 1)$-vector of cost coefficients, and since the rows of $\mathbf{B}^T$ are independent there is always a solution of these equations for $\mathbf{u}$ and $\mathbf{v}$ (see exercise 10.7).

For small "academic" examples a suitable circuit of basic variables can be found by inspection. In practice, for larger problems, a systematic search procedure is needed and one way to organise this is suggested by the technique used in the following two chapters where network flows are discussed.

The method developed in section 10.2 for solving transportation problems is sometimes called the *stepping-stone method.*

10.4

(i) The northwest corner solution is not necessarily the best initial *b.f.s.* to use, and any other *b.f.s.* with a lower cost would be preferable, although it would not necessarily result in fewer stages to obtain the optimum solution. One alternative which usually gives an improved initial *b.f.s.* (but sometimes a worse one!) is the matrix minimum method. Here, instead of starting with the northwest corner element of $\mathbf{C}$, we start with c_{st} where

$$c_{st} = \min_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,n}} c_{ij}.$$

This determines x_{st} and effectively reduces the problem by one row or column.

A compromise between this and the northwest corner method, to save repeatedly searching the whole matrix, is to choose the variable x_{ij} corresponding to the minimum cost coefficient in each row (or column) in turn instead of the northwest corner coefficient. For the example of section 10.2 two of these approaches yield the initial *b.f.s.s* below, where the integers in the bottom left-hand

corners indicate the order in which the basic variables are determined.

2 **2** 4	3 **2** 5	4	3	4
4	3	2 **0** 3	1 **4** 2	4
2	4 **4** 6	5 **4** 7	4	8
0 **1** 1	0	0	0	1
3	6	4	4	

matrix minimum method
cost = 58

2 **3** 1	3	4 **1** 6	3	4
4	3	2 **0** 3	1 **4** 2	4
2	4 **6** 4	**2** 7	4	8
0	0	0 **1** 5	0	1
3	6	4	4	

successive row minimum method
cost = 48

(ii) If **b** and **d** have integer elements then $\mathbf{x}_{opt}$ necessarily has integer elements, which is an important result for practical purposes (see also section 10.5). This follows directly from the method developed in section 10.2, which nowhere involves division. Notice however, that the method does not *require* that **b** and **d** have integer elements. The crucial aspect of transportation problems is that although feasible solutions **x** need not have integer elements when **b** and **d** have integer elements (e.g. $x_{ij} = b_j d_i / \Sigma_j b_j$), when this *is* the case any *b.f.s.* must have integer elements (see exercise 10.6).

(iii) The possibility of cycling (see section 4.7) would appear to be more serious for transportation problems, partly because *b.f.s.s* often have several basic variables with value zero, and partly because when all the numbers involved are integers the arithmetic

operations will be performed exactly. Nevertheless, transportation type problems which cycle are not expected to occur in practice so that the perturbation technique which follows is of mainly academic interest.

As with the basic simplex method, to prevent cycling it is sufficient to prevent degeneracy, and to prevent degeneracy it is sufficient to prevent ties between remaining column sums and remaining row sums. This can be done by replacing d_i by $d_i + \varepsilon$, $i = 1, 2, \ldots, m$, and b_n by $b_n + m\varepsilon$, for some positive but sufficiently small ε. As in section 4.7 a specific value need not be chosen for ε, and we just use the principle to decide which variables x_{ij} are basic with value zero. A different approach to transportation problems which uses graph theory may be found in {10}.

10.5 Assignment Problems

The transportation problem can be regarded as a problem in assigning the amounts of the commodity at each source to go to each destination, with a specific penalty, the cost, for each source and destination pair. The corresponding situation in which there is a benefit instead of a penalty clearly leads to a *l.p.p.* in which **A** has the same structure and a typical situation is that of personnel assignment. We shall distinguish three assignment problems:

(i) the simple assignment problem,
(ii) the optimum assignment problem,
(iii) the categorised optimum assignment problem.

The first two of these, although they are of transportation type, are even more specialised and may be solved by the special methods developed in chapter 12.

In the categorised optimum assignment problem, we may consider the situation of n categories of job with $b_1, b_2, \ldots, b_n$ vacancies respectively, and m categories of applicant with $d_1, d_2, \ldots, d_m$ persons respectively. For each category of applicant and each category of job, there is a rating which gives a numerical measure of the applicants' suitability for the jobs, and the problem is to decide how many persons from each category to assign to the various jobs so that the sum of the assignment ratings is maximised.

We may list the ratings r_{ij} in a rating matrix $\mathbf{R}$. Then with x_{ij} as the number of applicants in the i-th category assigned to the j-th job, the *l.p.p.* may be written

$$\textit{maximise} \quad \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij} x_{ij} \quad \textit{subject to} \tag{1}$$

$$\sum_{i=1}^{m} x_{ij} \geq b_j, \ \sum_{j=1}^{n} x_{ij} \leq d_i, \ x_{ij} \geq 0, \ i = 1, 2, \ldots, m, \ j = 1, 2, \ldots, n.$$

There would normally be the additional requirement that x_{ij} is an integer, but we know this will be the case for the optimum solution.

This is exactly the form of the transportation problem. If $\sum_{j=1}^{n} b_j \neq \sum_{i=1}^{m} d_i$ we introduce a fictitious category of person or job, and if we put $c_{ij} = -r_{ij}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, then the *l.p.p.* is exactly that of (2) or (3) of section 10.1.

Example

Three categories of applicant with 5, 8, 4 persons respectively, apply for five types of job with 4, 2, 1, 7, 3 vacancies respectively. The rating matrix is

$$\begin{pmatrix} 3 & 2 & 5 & 8 & 1 \\ 1 & 1 & 2 & 3 & 1 \\ 2 & 8 & 1 & 1 & 2 \end{pmatrix}.$$

Find the assignment which maximises the sum of the assigned ratings. For this example we find the initial *b.f.s.* not by the northwest corner method but by choosing minimum column elements. Thus $x_{11} = 4$, $x_{32} = 2$, $x_{13} = 1$, $x_{14} = 0$, $x_{24} = 7$, $x_{35} = 2$, $x_{25} = 1$, giving an initial cost of -59, that is a rating of 59.

u_i \ v_j	-3	-12	-5	-8	-6	
0	-3 / $4-\theta$	-2 / ✓ / -12	-5 / 1	-8 / $0+\theta$	-1 / ✓ / -6	5
5	-1 / / 2	-1 / ✓ / -7	-2 / / 0	-3 / $7-\theta$	-1 / $1+\theta$	8
4	-2 / θ / 1	-8 / 2	-1 / ✓ / -1	-1 / ✓ / -4	-2 / $2-\theta$	4
	4	2	1	7	3	b_j \ d_i

Value $= -(12 + 5 + 21 + 1 + 16 + 4)$,
i.e. *rating* $= 59$,
min $c'_{ij} = c'_{21} = c'_{31} = -3$,
choosing c'_{31} because $c_{31} < c_{21}$
leads to $\theta = 2$.
$c'_{31}\theta = -6$.

$u_i \backslash v_j$	-3	-9	-5	-8	-6	
0	$^{-3}$ $2-\theta$	$^{-2}$ $\surd$ $_{-9}$	$^{-5}$ 1	$^{-8}$ $2+\theta$	$^{-1}$ $\surd$ $_{-6}$	5
5	$^{-1}$ θ $_{2}$	$^{-1}$ $\surd$ $_{-4}$	$^{-2}$ $_{0}$	$^{-3}$ $5-\theta$	$^{-1}$ 3	8
1	$^{-2}$ 2	$^{-8}$ 2	$^{-1}$ $\surd$ $_{-4}$	$^{-1}$ $\surd$ $_{-7}$	$^{-2}$ $\surd$ $_{-5}$	4
	4	2	1	7	3	$b_j \backslash d_i$

Value $= -(6 + 5 + 16 + 15 + 3 + 4 + 16)$,
i.e. *rating* $= 65$,
min $c'_{ij} = c'_{21} = -3$,
$\theta = 2$.
$c'_{23}\theta = -2$.

$u_i \backslash v_j$	-6	-12	-5	-8	-6	
0	$^{-3}$ $\surd$ $_{-6}$	$^{-2}$ $\surd$ $_{-12}$	$^{-5}$ $1-\theta$	$^{-8}$ $4+\theta$	$^{-1}$ $\surd$ $_{-6}$	5
5	$^{-1}$ 2	$^{-1}$ $\surd$ $_{-7}$	$^{-2}$ θ $_{0}$	$^{-3}$ $3-\theta$	$^{-1}$ 3	8
4	$^{-2}$ 2	$^{-8}$ 2	$^{-1}$ $\surd$ $_{-1}$	$^{-1}$ $\surd$ $_{-4}$	$^{-2}$ $\surd$ $_{-2}$	4
	4	2	1	7	3	$b_j \backslash d_i$

Value $= -(5 + 32 + 2 + 9 + 3 + 4 + 16)$,
i.e. *rating* $= 71$,
min $c'_{ij} = c'_{23} = -2$,
$\theta = 1$.
$c'_{23}\theta = -2$.

$u_i \backslash v_j$	-6	-12	-7	-8	-6	
0	$^{-3}$ $\surd$ $_{-6}$	$^{-2}$ $\surd$ $_{-12}$	$^{-5}$ $\surd$ $_{-7}$	$^{-8}$ 5	$^{-1}$ $\surd$ $_{-6}$	5
5	$^{-1}$ 2	$^{-1}$ $\surd$ $_{-7}$	$^{-2}$ 1	$^{-3}$ 2	$^{-1}$ 3	8
4	$^{-2}$ 2	$^{-8}$ 2	$^{-1}$ $\surd$ $_{-3}$	$^{-1}$ $\surd$ $_{-4}$	$^{-2}$ $\surd$ $_{-2}$	4
	4	2	1	7	3	$b_j \backslash d_i$

Value $= -(40 + 2 + 2 + 6 + 3 + 4 + 16)$,
i.e. *rating* $= 73$,
all $c'_{ij} \geq 0$,
hence current b.f.s. is optimum.

The dual variables **u**, **v** satisfy the constraints, and the value of the dual solution is

$$40 + 16 - 24 - 24 - 7 - 56 - 18 = -73.$$

Thus the maximum possible overall rating is 73, and an assignment that gives this rating is: 2 persons from group 2 and 2 from group 3 do job 1, 2 persons from group 3 do job 2, 1 person from group

2 does job 3, 5 persons from group 1 and 2 from group 2 do job 4, and 3 persons from group 2 do job 5.

Instead of converting ratings to costs and minimising, we could maximise the sum of the assigned ratings directly, in which case the optimality criterion would be $c'_{ij} \leq 0$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$; it is a matter of personal opinion which is least confusing.

10.6 The Caterer's Problem

The transportation problem appears in a number of situations. The trans-shipment problem is one (see exercise 10.14), and the contract award problem (see exercise 1.4) is another. A particularly ingenious application involves a caterer who requires clean table-cloths for dinner-parties on successive days. The table-cloths can be purchased for c_1, cleaned overnight for c_2 or cleaned over a period of p days for c_3 (i.e. used on day j and ready again on day $(j+p+1)$, where $c_1 > c_2 > c_3$. Assuming that any number of table-cloths may be purchased and any number cleaned by either laundry service on any day (and assuming that they are unerringly soiled by the diners), how should the caterer arrange for the daily supply of clean table-cloths so that the cost of providing them is minimised?

Suppose there are n dinner-parties, one on each of n successive days, the j-th one requiring b_j table-cloths, $j = 1, 2, \ldots, n$; these are the destinations. The sources are the supplier, whom we assume has $d_{n+1} = \Sigma_{j=1}^{n} b_j$ available for purchase, and the n baskets of soiled table-cloths at the end of each party, the i-th one containing d_i, where $d_i = b_i$, $i = 1, 2, \ldots, n$. It is convenient to introduce an aftermath destination which requires b_{n+1} table-cloths, where $b_{n+1} = d_{n+1}$. If we denote by x_{ij} the number of table-cloths used on the j-th day from the i-th source, then the constraints are

$\Sigma_{i=1}^{n+1} x_{ij} = b_j$ (= total number used on j-th day), $j = 1, 2, \ldots, n+1$,

and $\Sigma_{j=1}^{n+1} x_{ij} = d_i$(= b_i = total number used from i-th source on all days, plus the number going to the final destination from the i-th source), $i = 1, 2, \ldots, n+1$,

and $x_{ij} \geq 0$, $i, j = 1, 2, \ldots, n+1$.

The cost coefficients are:

$$c_{ij} = \begin{cases} 0, \; j = n+1, \; i = 1, 2, \ldots, n \\ c_1, \; i = n+1, \; j = 1, 2, \ldots, n \\ c_2, \; i = 1, 2, \ldots, n-1, \; j = i+1, \; i+2, \ldots, \min(i+p, n) \\ c_3, \; i = 1, 2, \ldots, n-p-1, j = i+p+1, \ldots, n \\ \infty, \; i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, i. \end{cases} \qquad (ER)$$

Exercises 10

1. Interpret the dual of the transportation problem in canonical primal form, in terms of a haulage company who offer to buy the commodity at the sources where it is manufactured and sell it back to the manufacturer at the destinations.
2. Solve the transportation problem in which $m = 3$, $n = 5$, $\mathbf{d}^T = (4, 5, 6)$, $\mathbf{b}^T = (2, 2, 3, 4, 4)$ and
$$\mathbf{C} = \begin{pmatrix} 3 & 6 & 3 & 1 & 1 \\ 2 & 4 & 3 & 2 & 7 \\ 4 & 1 & 2 & 1 & 2 \end{pmatrix}.$$
3. Obtain a sufficient condition for the optimum solution of the transportation problem to be unique (see exercise 3.6). Obtain a different optimum transportation scheme for the examples of exercise 10.2 and of section 10.2.
4. Prove that the matrix $\mathbf{A}$ of a transportation problem has rank $(m + n - 1)$.
5. For $m = 3$, $n = 5$ say, and supposing that the first row of the matrix $\mathbf{A}$ of a transportation problem is removed, choose any $7\ (= m + n - 1)$ independent columns and show that they can be rearranged by row and column interchanges to form an upper triangular matrix with unit diagonal. What is the implication of this result? (The result holds in general.)
6. A *companion* exercise to 10.5 which requires a knowledge of determinants: Prove that all minors of $\mathbf{A}$ (determinants of square submatrices of $\mathbf{A}$) have value -1, 0, or $+1$. Hence explain why the inverse of any non-singular $(m + n - 1) \times (m + n - 1)$ submatrix of $\mathbf{A}$ has only integer elements, and hence why, if d_i, b_j are all integers, then the optimum solution of a transportation problem has only integer elements.
7. Let $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{2k}$ be the columns of a transportation matrix $\mathbf{A}$ corresponding to the variables x_{ij} in a *θ-circuit*. Show that the vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{2k}$ are linearly dependent and that $\Sigma_{i=0}^{2k} \alpha_i \mathbf{z}_i = \mathbf{0}$ with each $\alpha_i = \pm 1$. Explain why any $(2k - 1)$ of the vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{2k}$ are linearly independent. Hence prove that at every stage of the method of section 10.2 the columns of $\mathbf{A}$ corresponding to basic variables are linearly independent.
8. Solve the example of section 10.2 starting with the initial *b.f.s.* obtained by choosing (i) the matrix minimum method, and (ii) the successive column minimum method.

9. a) Show that during the solution of a transportation problem the vectors **u**, **v**, **x** at every stage satisfy

$$\Sigma_{ij} c'_{ij}\, x_{ij} = \Sigma_i u_i d_i + \Sigma_j v_j b_j.$$

b) Explain whether, instead of putting $u_1 = 0$ at each stage, any element of **u** or **v** can be given an arbitrary value.

10. Solve the categorised optimum assignment problem in which five categories of person with 5, 6, 3, 1, 3 persons in each category respectively, apply for three categories of job with 7,6,4 vacancies respectively and the rating matrix is

$$R = \begin{pmatrix} 3 & 2 & 7 \\ 1 & 4 & 2 \\ 6 & 1 & 5 \\ 3 & 2 & 3 \\ 2 & 3 & 2 \end{pmatrix}.$$

11. A personnel officer, having solved a categorised assignment problem, decides to revise the rating matrix **R**. The new ratings $\tilde{r}_{ij}$ are given by $\tilde{r}_{ij} = \alpha r_{ij} + \beta_j$, for some constant α and constants $\beta_1, \beta_2, \ldots, \beta_n$. Devise an efficient way (!) of obtaining the revised optimum solution.

12. Solve the categorised optimum assignment problem of section 10.5 starting with the initial *b.f.s.* given by
    (i) the northwest corner method,
    (ii) the matrix minimum method, and
    (iii) the successive row minimum method.

13. Solve the caterer's problem in which there are four dinner parties on successive days requiring respectively 20, 27, 38, 28 table-cloths which cost 9 units to buy, 4 units to clean overnight or 2 units to clean by the day after next.

14. One version of the trans-shipment problem is the transportation problem in which there are intermediate junctions where loads of the commodity can be divided and reassembled and which have a maximum capacity. Ignoring any costs arising from the redistribution, and assuming each part-route has a transportation cost per unit of commodity, express the trans-shipment problem as a *l.p.p.*

15. Prove that in the optimum solution **x** of a transportation type problem at least one variable x_{ij} is equal to b_i or d_j.
(Only true when optimum solution is unique.)

NOTES

CHAPTER 11

NETWORK FLOWS

11.1

The solution of a transportation problem can be thought of as defining a flow of the commodity from the set of sources to the set of destinations along the routes connecting them. A time period was never mentioned in connection with transportation problems and so the resulting solution can refer to a single transportation task, an annual programme, or a weekly, or a daily one. Alternatively, all amounts of commodity can be interpreted as rates of flow, so that for example, $b_1 = 2b_2$ means that whatever quantity of the commodity is delivered to the second destination, twice that quantity is delivered to the first, and so on. If we imagine a more complicated network of routes connecting sources and destinations, with intermediate junctions, and instead of a unit cost for each part of each route there is a maximum capacity, then the problem of determining the maximum possible flow is clearly a *l.p.p.* (see exercise 11.1). Instead of developing a special version of the simplex method to solve such problems we develop an independent method. We restrict our attention to networks with a single source s and a single destination s', which in this context is called a *sink*, but see exercise 11.5. The method, or algorithm, we develop can be used to solve network flow problems with integer or with arbitrary capacities but as we shall use it in chapter 12 for assignment problems we will only consider problems in which the capacities are integers.

The points of a network, the source(s), sink(s) and intermediate points, are called *nodes* and the connecting routes are called *edges.* The nodes are denoted by $x_1, x_2, \ldots, x_n$, an edge by (x_i,x_j) and the whole set of nodes by N. A *capacity function* k assigns to each edge of N a non-negative integer $k(x_i,x_j)$ which is the maximum flow from x_i to x_j that the edge (x_i,x_j) can support. Capacities may be symmetric ($k(x_i,x_j) = k(x_j,x_i)$) or unsymmetric ($k(x_i,x_j) \neq k(x_j,x_i)$) and $k(x_i,x_i) = 0$. A *capacitated network* (N,k) is a network N together with the associated capacity function k.

A flow in a capacitated network (N,k) is a function which assigns to each edge (x_i,x_j) a number $f(x_i,x_j)$ and which satisfies

$$f(x_i,x_j) = -f(x_j,x_i), \tag{1}$$

$$\text{and} \quad f(x_i,x_j) \leq k(x_i,x_j). \tag{2}$$

A flow function is essentially just a list of flows, and will be integer valued as far as we are concerned. Notice that (1) implies that $f(x_i,x_i) = 0$ and introduces the convention of *nett* flows. If there is a flow α from x_i to x_j and β from x_j to x_i this is the same as a flow $\alpha - \beta$ from x_i to x_j.

For the two functions k and f defined on a network N we shall use the notation $k(A,B)$ and $f(A,B)$ to denote $\Sigma_{\substack{x_i \in A \\ x_j \in B}} k(x_i,x_j)$ and $\Sigma_{\substack{x_i \in A \\ x_j \in B}} f(x_i,x_j)$, where A and B are *any* subsets of N. The properties (1) and (2) imply that

$$f(A,A) = 0 \quad \text{and} \quad f(A,B) \leq k(A,B).$$

Also, for any distinct subsets A and B of N, and any subset C of N,

$$f(A \cup B, C) = f(A,C) + f(B,C),$$

$$\text{and} \quad f(C, A \cup B) = f(C,A) + f(C,B) \tag{3}$$

and the same is true for k.

We formally define a source s and a sink s' for a flow f in a network by saying a node s is a source for f if

$$f(s,N) > 0 \quad \text{and} \quad f(s,x_i) \geq 0, \ x_i \in N;$$

and a node s' is a sink for f if

$$f(N,s') > 0 \quad \text{and} \quad f(x_i,s') \geq 0, \ x_i \in N.$$

The second condition in both cases is to avoid any possible complications with unproductive circular flows, for example from s to x_1 to x_2 to x_3 to s, and from now on N will be used to denote all the nodes $x_1, x_2, \ldots, x_n$ together with s and s'.

We can now define the problem as follows: given a capacitated network (N,k) with a single source s and a single sink s', find a flow f whose value $f(s,N)$ is a maximum.

For any (finite) network a *maximum flow* exists, and its value is at most $k(s,N)$ (ER).

11.2

To develop a method for finding a maximum flow we need another concept, that of a *cut* in a network.

A *cut* (S,S') in a capacitated network (N,k) with a single source s and a single sink s' is a division of the nodes of N into two disjoint subsets S and S' which satisfy $S \cup S' = N$, $s \in S$, $s' \in S'$.

The *capacity of a cut* is defined as $k(S,S')$, and a cut which has the minimum possible value is called a *minimum cut*.

The connection between flows and cuts begins with the following observation: if (S,S') is *any* cut in a capacitated network (N,k) with a single source s and a single sink s', and f is *any* flow, then the value of the flow $f(s,N)$ is at most the capacity of the cut $k(S,S')$. If $f(s,N) = k(S,S')$ then f is a *maximum flow* and (S,S') is a *minimum cut*.

This result is easily established using (1) (2) and (3) of section 11.1. As $f(x_i,N) = 0$ for $x_i \neq s', s$,

$$\begin{aligned} f(s,N) &= f(s,N) + f(X_s,N), \quad \text{where } x_i \in X_s \text{ if } x_i \in S \text{ and } x_i \neq s, \\ &= f(S,N) = f(S,S \cup S') \\ &= f(S,S) + f(S,S') = f(S,S') \leq k(S,S'). \end{aligned}$$

If a particular flow f_0 and a particular cut (S_0,S_0') satisfy $f_0(s,N) = k(S_0,S_0')$ then $f(s,N) \leq k(S_0,S_0')$ for any flow, and hence f_0 is a maximum flow, and similarly (S_0,S_0') is a minimum cut.

The correspondence between cuts and flows and primal and dual *l.p.p.s* is already apparent and is emphasised by the next theorem.

Theorem 13. The Maximum Flow-Minimum Cut Theorem

For any capacitated network with a single source and a single sink the value of a maximum flow is equal to the value of a minimum cut∎

Let $\tilde{f}$ be a maximum flow. We say an edge (x_i,x_j) is *saturated* by $\tilde{f}$ if

$$\tilde{f}(x_i,x_j) = k(x_i,x_j).$$

A *path* is a sequence of edges of N connecting distinct nodes of N, and an *unsaturated path* is a path all of whose edges are unsaturated.

Thus a path from s to x_t can be denoted by

$$P = \{s, x_{i_1}, x_{i_2}, \ldots, x_{i_p}, x_t\}.$$

The edges of P are (s,x_{i_1}), (x_{i_1},x_{i_2}), $\ldots$, (x_{i_p},x_t), and if P is unsaturated then for any edge (x_i,x_j) of P

$$\tilde{f}(x_i,x_j) < k(x_i,x_j).$$

Now we define sets S and S' by saying $s \in S$, and $x_i \in S$ if $x_i \in N$ and there is an unsaturated path from s to x_i; $x_i \in S'$ if $x_i \in N$ and $x_i \notin S$.

To show that (S,S') is a cut we need to show that $s' \notin S$. So suppose that $s' \in S$; then there is an unsaturated path P from s to s'. Let

$$\delta = \min_{(x_i,x_j)\in P} (k(x_i,x_j) - \tilde{f}(x_i,x_j))$$

and define a flow f by

$$f(x_i,x_j) = \tilde{f}(x_i,x_j) \quad \text{for} \quad (x_i,x_j) \notin P \quad \text{and}$$
$$f(x_i,x_j) = \tilde{f}(x_i,x_j) + \delta \quad \text{for} \quad (x_i,x_j) \in P.$$

From the definition of P, $\delta > 0$ and the flow f satisfies

$$f(x_i,x_j) \leq k(x_i,x_j) \quad \text{for all edges of } N,$$

but the value of the flow f is

$$f(s,N) = \tilde{f}(s,N) + \delta,$$

which contradicts the flow $\tilde{f}$ being maximum. Thus $s' \in S'$ and (S,S') is a cut.

Now we already know that

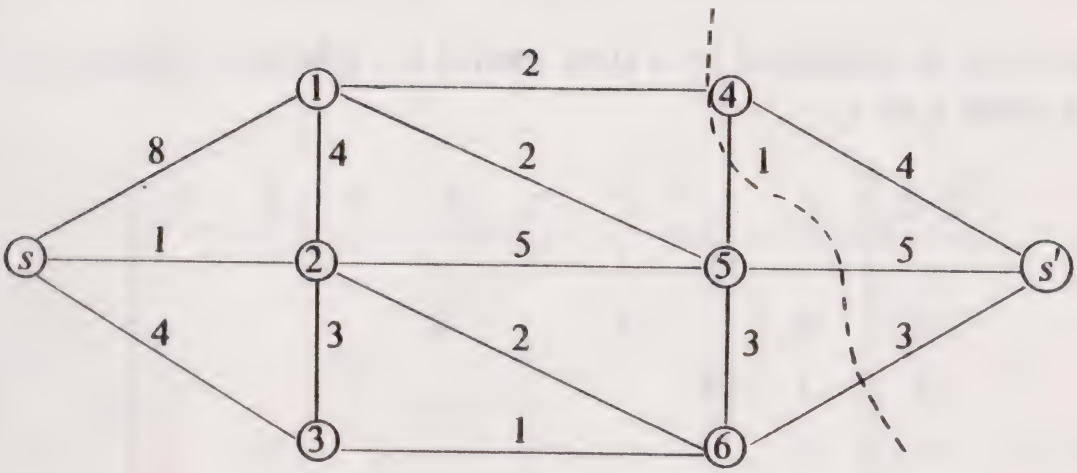
$$\tilde{f}(s,N) = \tilde{f}(S,S') \leq k(S,S').$$

Hence, if $\tilde{f}(s,N) < k(S,S')$ then $\tilde{f}(x_i,x_j) < k(x_i,x_j)$ for some edge (x_i,x_j) with $x_i \in S$ and $x_j \in S'$, and so the unsaturated path from s to x_i can be extended to x_j which contradicts the definition of S and S'. Thus $\tilde{f}(s,N) = k(S,S')$ ■

Theorem 13 is clearly the counterpart for networks of the duality theorem for *l.p.p.s*, and like the duality theorem, it gives a means of testing whether a flow is maximum. It also suggests a method of constructing a maximum flow by listing, for any current flow, the set of nodes that can be reached from s by unsaturated paths. If this set contains s', we can improve the current flow and repeat; if this set does not contain s' we can verify that the flow is a maximum flow using the cut defined by the set. A systematic way of implementing this method is described by an example.

11.3

Find a maximum flow in the network:



,

where the edges have the symmetric capacities indicated.

The network can be described by its capacity matrix **K**, where k_{ij} denotes the capacity of the edge (x_i, x_j) and where entries are made only of elements which have a corresponding edge in the network. The initial capacity matrix, referring to the network with no flow defined, we denote by $\mathbf{K}_0$.

$\mathbf{K}_0 =$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		8	1	4				
1	8		4		2	2		
2	1	4		3		5	2	
3	4		3				1	
4		2				1		4
5		2	5		1		3	5
6			2	1		3		3
s'					4	5	3	

We can see by inspection that a flow f_1 can be imposed consisting of

2 units from s to x_1 to x_4 to s',

1 unit from s to x_2 to x_5 to s',

1 unit from s to x_3 to x_6 to s',

4 units from s to x_1 to x_2 to x_5 to s'.

This flow can be described by a flow matrix $\mathbf{F}_1$, where $(\mathbf{F}_1)_{ij}$ denotes the flow from x_i to x_j.

$\mathbf{F}_1 =$

$x_i \diagdown x_j$	s	1	2	3	4	5	6	s'
s		6	1	1				
1	−6		4		2			
2	−1	−4				5		
3	−1						1	
4		−2						2
5			−5					5
6				−1				1
s'					−2	−5	−1	

Since the edges of N have symmetric capacities, $\mathbf{K}_0$ is symmetric. The matrix $\mathbf{F}_1$ is skew-symmetric. The zero elements of $\mathbf{F}_1$ and the elements for which there is no corresponding edge have been omitted.

The flow of 1 unit from s to x_3 means, for example, that the edge (s,x_3) now has capacity $4 - 1 = 3$ units and the edge (x_3,s) now has capacity $4-(-1) = 5$ units. Overall, the capacity of N with the flow f_1 is given by $\mathbf{K}_1 = \mathbf{K}(f_1) = \mathbf{K}_0 - \mathbf{F}_1$.

$\mathbf{K}_1 =$

$x_i \diagdown x_j$	s	1	2	3	4	5	6	s'
s		2	0	3				
1	14		0		0	2		
2	2	8		3		0	2	
3	5		3				0	
4		4				1		2
5		2	10		1		3	0
6			2	2		3		2
s'					6	10	4	

Using $\mathbf{K}_1$ we now search for an unsaturated path from s to s'. The edge from x_i to x_j is unsaturated if $(\mathbf{K}_1)_{ij} > 0$, so searching the *s-row* of $\mathbf{K}_1$ we find unsaturated edges (s,x_1), (s,x_3), which we can denote by

$$s \rightarrow \begin{cases} x_1 \\ x_3 . \end{cases}$$

Now searching the x_1-row of $\mathbf{K}_1$ we find that the edges (x_1,s) and (x_1,x_5) are unsaturated, but as s has appeared before in this search, the edge (x_1,s) can be ignored. Thus we have

$$x_1 \rightarrow \{x_5 ,$$

and the same procedure for x_3 gives

$$x_3 \rightarrow \{x_2 ,$$

where the edge (x_3,s) has been ignored because s has already appeared.

Combining both stages gives

$$s \rightarrow \begin{cases} x_1 \rightarrow \{x_5 \\ x_3 \rightarrow \{x_2 , \end{cases} \tag{1}$$

and continuing with the x_5 and x_2-rows of $\mathbf{K}_1$ gives

$$\begin{aligned} x_5 &\rightarrow \begin{cases} x_4 \\ x_6 \end{cases} \\ x_2 &\rightarrow \{X \ , \end{aligned} \tag{2}$$

where X denotes that no further progress can be made since all unsaturated edges from x_2, (x_2,s), (x_2,x_1), (x_2,x_3), (x_2,x_6), lead to nodes which already appear in the *tree* (1) and (2).

The whole *tree* so far is

$$s \rightarrow \begin{cases} x_1 \rightarrow \{x_5 \rightarrow \begin{cases} x_4 \\ x_6 \end{cases} \\ x_3 \rightarrow \{x_2 \rightarrow X \ . \end{cases} \tag{3}$$

The x_4-row of $\mathbf{K}_1$ gives

$$x_4 \rightarrow \{s' ,$$

so we have an unsaturated path P_1,

$$P_1 = \{s, x_1, x_5, x_4, s'\}.$$

The search procedure finds systematically an unsaturated path from s to s' if there is one. It does not find all unsaturated paths nor the unsaturated path with greatest capacity. We would find a different path if we considered x_5 in the tree (1) before considering x_2, or if we considered x_6 in the tree (2) before considering x_4.

The minimum capacity of edges on the path P_1 is

$$min\ \{(\mathbf{K}_1)_{S1}, (\mathbf{K}_1)_{15}, (\mathbf{K}_1)_{54}, (\mathbf{K}_1)_{4S}\} = min\ \{2,2,1,2\} = 1 .$$

Thus the flow δf_1 described by $\delta\mathbf{F}_1$ can be added to f_1,

$\delta\mathbf{F}_1 =$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		1						
1	−1					1		
2								
3								
4						−1		1
5		−1			1			
6								
s'					−1			

giving a total current flow f_2 described by $\mathbf{F}_2$, where $\mathbf{F}_2 = \mathbf{F}_1 + \delta\mathbf{F}_1$.

$\mathbf{F}_2 =$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		7	1	1				
1	−7		4		2	1		
2	−1	−4				5		
3	−1						1	
4		−2				−1		3
5		−1	−5		1			5
6				−1				1
s'					−3	−5	−1	

We can now describe the current capacity of N with f_2 by

$$\mathbf{K}_2 = \mathbf{K}_1 - \delta\mathbf{F}_1 = \mathbf{K}_0 - \mathbf{F}_2.$$

$\mathbf{K}_2 =$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		1	0	3				
1	15		0		0	1		
2	2	8		3		0	2	
3	5		3				0	
4		4				2		1
5		3	10		0		3	0
6			2	2		3		2
s'					7	10	4	

The search tree for $\mathbf{K}_2$ is

$$s \rightarrow \begin{cases} x_1 \rightarrow \{x_5 \rightarrow \{x_6 \rightarrow s' \\ x_3 \rightarrow \{x_2 \rightarrow \{X \end{cases}$$

and a further flow of 1 unit

from $\quad s$ to x_1 to x_5 to x_6 to s'

can be added to f_2. To save writing out a further stage we can observe here that there is still another unsaturated path,

from $\quad s$ to x_3 to x_2 to x_6 to s',

which can accomodate a flow of 1 unit (the capacity of (x_6, s') has just been reduced to 1 unit). These two additional flows constitute δf_2.

$\delta\mathbf{F}_2 =$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		1		1				
1	–1					1		
2				–1			1	
3	–1		1					
4								
5		–1					1	
6			–1			–1		2
s'							–2	

$\mathbf{F}_3 = \mathbf{F}_2 + \delta\mathbf{F}_2 =$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		8	1	2				
1	–8		4		2	2		
2	–1	–4		–1		5	1	
3	–2		1				1	
4		–2				–1		3
5		–2	–5		1		1	5
6			–1	–1		–1		3
s'					–3	–5	–3	

$$\mathbf{K}_3 = \mathbf{K}_2 - \delta\mathbf{F}_2 = \mathbf{K}_0 - \mathbf{F}_3 =$$

$x_i \backslash x_j$	s	1	2	3	4	5	6	s'
s		0	0	2				
1	16		0		0	0		
2	2	8		4		0	1	
3	6		2				0	
4		4				2		1
5		4	10		0		2	0
6			3	2		4		0
s'					7	10	6	

The search tree for $\mathbf{K}_3$ is

$$s \to \{x_3 \to \{x_2 \to \begin{cases} x_1 \to \{X \\ x_6 \to \{x_5 \to \{X, \end{cases} \tag{4}$$

which shows that $f_3 = f_2 + \delta f_2$ is a maximum flow.

A minimum cut is given (from (4)) by

$$S = \{s, x_1, x_2, x_3, x_5, x_6\}, \quad S' = \{s', x_4\}.$$

The value of f_3 is the sum of the elements in the first row of $\mathbf{F}_3$, which is

$$8 + 1 + 2 = 11.$$

The capacity of the cut which is indicated by the dotted line on the diagram on page 145 is the sum of the capacities of edges which cross the cut and is

$$2 + 1 + 5 + 3 = 11.$$

11.4

The method of the previous section is rather tedious for small networks, particularly those which can easily be described in a two-dimensional diagram and which can usually be solved by inspection, using the principles of the method but without writing out the capacity and flow matrices $\mathbf{K}$ and $\mathbf{F}$. The procedure at each stage involves a more exhaustive and less precisely defined search than that needed at each stage of the simplex method, and this aspect is typical of algorithms for integer linear programming problems.

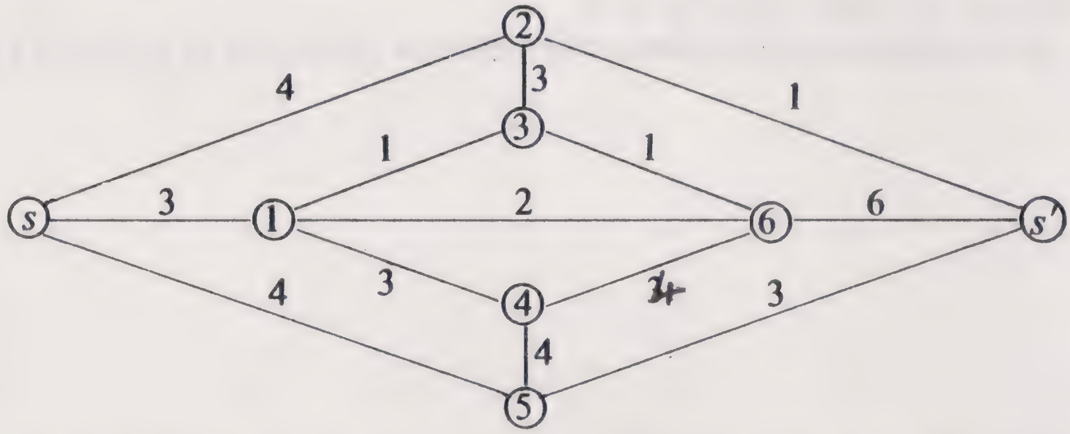
To implement the method in practice one has various options, most of which make an insignificant difference to the efficiency. The most natural approach is to store (row-wise) only the elements of $\mathbf{K}$ and

F corresponding to edges that are present in N, and to store $(\mathbf{K})_{ij}$ and $(\mathbf{F})_{ij}$ in consecutive locations of a one-dimensional array. The essential point is that both **K** and **F** are sparse, often very sparse, and the elements which may be non-zero are defined by the network and do not change during the algorithm. It is more convenient to store $\mathbf{K}_i$ and overwrite at each stage, but it makes no difference whether we also store $\mathbf{K}_0$ or $\mathbf{F}_i$.

A comprehensive treatment of network problems is given in {11}.

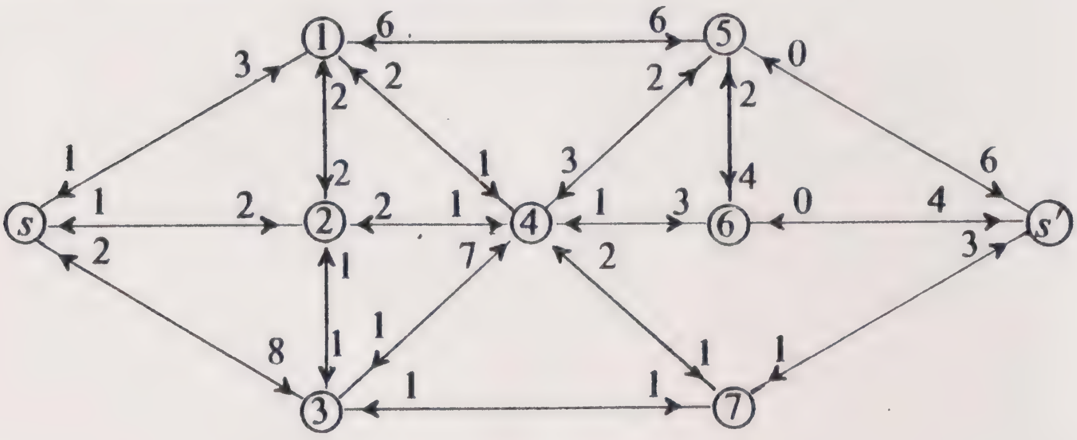
Exercises 11

1. For a capacitated network (N,k) with a single source s and a single sink s' write the problem of finding a maximum flow as a *l.p.p.* in standard (dual) form.
2. Find a maximum flow in the network below, where the capacities indicated are symmetric.



Find by inspection an alternative maximum flow and the corresponding minimum cut.

3. Find a maximum flow and a minimum cut for the network below, where the capacities are as indicated by the arrows.



4. Denoting the s- and s'-rows and columns of capacity and flow matrices by the suffices 0 and $n+1$ respectively, show that

   (i) for any flow matrix $\mathbf{F}_t$ and any i, $i = 0, 1, \ldots, n+1$,

   $$\Sigma^{i}_{j=0} (\mathbf{F}_t)_{i-j,j} = 0, \text{ and}$$

   (ii) for any capacity matrix $\mathbf{K}_t$ and any i, $i = 0, 1, \ldots, n+1$,

   $$\Sigma^{i}_{j=0} (\mathbf{K}_t)_{i-j,j} = \Sigma^{i}_{j=0} (\mathbf{K}_0)_{i-j,j}.$$

5. Suppose a capacitated network (N,k) has t sources $s_1, s_2, \ldots, s_t$, t' sinks $s'_1, s'_2, \ldots, s'_{t'}$, and n other nodes $x_1, x_2, \ldots, x_n$, and it is required to find a maximum flow from the set of sources to the set of sinks. Devise a capacitated network with a single source s and a single sink s' and $(n+t+t')$ other nodes whose maximum flow (or flows) provides a solution to the given problem.

NOTES

CHAPTER 12

ASSIGNMENT PROBLEMS: THE MARRIAGE PROBLEM

12.1

In this chapter we use an interesting and elegant network to solve the simple assignment problem and the optimum assignment problem mentioned in chapter 10. We describe each problem in terms of one particular situation, but many other situations clearly lead to the same mathematical problems. The essential feature is the requirement to match or assign the members of one set to the members of another, either so that some common criterion is satisfied or so that some quantitative measure of the success of the matching is optimised.

The Simple Assignment Problem

Suppose there are m individuals $I_1, I_2, \ldots, I_m$ to be assigned to n jobs $J_1, J_2, \ldots, J_n$. Each individual may only be assigned to those jobs for which he or she is qualified. The problem is to assign as many individuals as possible.

The situation may be described by the $m \times n$ *qualification matrix* **Q** in which

$$q_{ij} = 1 \quad \text{if } I_i \text{ is qualified for } J_j, \text{ and}$$
$$q_{ij} = 0 \quad \text{if } I_i \text{ is not qualified for } J_j.$$

The values 0 and 1 are unimportant; they are used here just as two different symbols. The problem may now be restated as:

given a qualification matrix **Q** find as many distinct 1's as possible such that no two of them are in the same row or column.

If $m \neq n$ we can introduce fictitious persons or jobs with no qualifications, i.e. rows or columns of **Q** with all entries zero. It is convenient to assume that this has been done so that $m = n$, but it is not necessary in practice and the method to be developed does not require it.

Assuming that the jobs are desired by the individuals and are not some form of punishment, then we have a convenient and benevolent view of the situation if we try to find every individual a job he

or she is qualified for. Clearly a necessary condition for all individuals to be assigned is that any set of p of the individuals must between them be qualified for at least p distinct jobs.

For example, with $m = n = 4$ and

$$\mathbf{Q} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \tag{1}$$

every individual is qualified for at least one job and for every job there is a qualified individual, but not all individuals can be assigned; I_1, I_2, I_4 are together qualified for only J_2 and J_4.

What is far less obvious is the converse result, that if, for $p = 1, 2, \ldots, n$, *any* p individuals are together qualified for at least p distinct jobs, then all n individuals may be assigned. This is the central result for the simple assignment problem, and we prove it using an *assignment network*.

An assignment network (N,k) has $(2n+2)$ nodes, one each for $I_1, I_2, \ldots, I_n$, $J_1, J_2, \ldots, J_n$ together with a source s and a sink s'. Its capacity function k is defined as follows:

$$\begin{aligned} k(s,I_i) &= 1, \ i = 1, 2, \ldots, n, \\ k(J_j,s') &= 1, \ j = 1, 2, \ldots, n, \\ k(I_i,J_j) &= K \ \text{ if } I_i \text{ is qualified for } J_j, \end{aligned} \tag{2}$$

where K is some large integer which need not be specified;

all other capacities are zero.

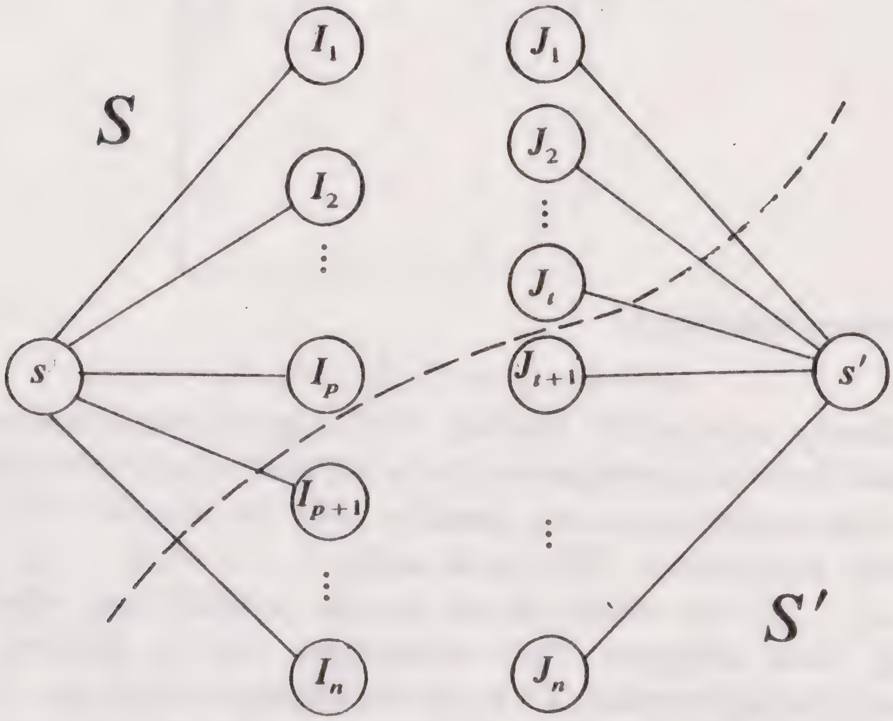
The maximum flow is n (or m if we had $m \neq n$ and $m < n$) because $k(s,N) = n$, and if f is any flow in (N,k) we define an assignment by saying I_i is assigned to J_j if $f(I_i,J_j) = 1$. If a flow has value n, then all n individuals are assigned. Since $f(s,N) = n = f(N,s')$ and $k(J_j,I_i) = 0$, the n unit flows from s to I_1, I_2, $\ldots$, I_n must continue to distinct job nodes.

12.2 Theorem 14

If, for $p = 1, 2, \ldots, n$, any p individuals are together qualified for at least p distinct jobs then all n individuals may be assigned∎

We prove that if all n individuals cannot be assigned then there must be a set of p individuals together qualified for less than p distinct jobs. So, suppose any maximum flow f in the assignment network has value less than n, and let (S,S') be a minimum cut. We may assume *w.l.o.g.* that

$$S = \{s, I_1, I_2, \ldots, I_p, J_1, J_2, \ldots; J_t\},$$
$$S' = \{s', I_{p+1}, I_{p+2}, \ldots, I_n, J_{t+1}, J_{t+2}, \ldots, J_n\}.$$



(1)

For $i \leq p$ and $j > t$, I_i is not qualified for J_j, i.e. $k(I_i,J_j) = 0$ for $i \leq p$ and $j > t$, because if I_i is qualified for J_j then

$$k(S,S') \geq k(I_i,J_j) = K > n \geq f(S,S') \tag{2}$$

and the same is true for $i > p$ and $J \leq t$.

Therefore

$$n > k(S,S') = \Sigma_{i>p}\, k(s,I_i) + \Sigma_{j \leq t}\, k(J_j,s') = (n-p) + t. \tag{3}$$

Hence $p > t$, and we have p individuals $I_1, I_2, \ldots, I_p$ qualified for fewer than p jobs, $J_1, J_2, \ldots, J_t$ ■

Thus to solve a simple assignment problem we find a maximum flow in the assignment network. If the flow has value n all individuals are assigned; if this cannot be done, the minimum cut provides a set of individuals and jobs which proves that it cannot be done and which indicates the number of individuals that cannot be assigned.

The assignment network has a particular form which enables us to avoid the $(2n+2) \times (2n+2)$ capacity and flow matrices of chapter 11, and to use instead just the assignment matrix **Q**.

12.3

We describe the method for finding a maximum flow in an assignment network using a simple problem with $m = n = 4$ and then examine a less trivial problem in the next section.

Suppose the qualification matrix is

$$
\begin{array}{c|cccc|}
I_i \backslash J_j & J_1 & J_2 & J_3 & J_4 \\
\hline
I_1 & 1 & & 1 & \\
I_2 & & 1 & 1 & \\
I_3 & 1 & 1 & & 1 \\
I_4 & 1 & & & \\
\hline
\end{array}
\quad . \qquad (1)
$$

An obvious solution is

$$I_1 \text{ to } J_3,\ I_2 \text{ to } J_2,\ I_3 \text{ to } J_4,\ I_4 \text{ to } J_1 .$$

It is clearly worthwhile starting with a good *initial* assignment, i.e. an *initial* flow in the assignment network with a high value (see exercise 12.8). For convenience we consider only the *simplest* way to obtain an initial assignment. This is to assign I_i, $i = 1, 2, \ldots, n$, to the first of $J_1, J_2, \ldots, J_n$ for which he or she is qualified and which has not already been assigned. This assignment can be denoted easily by replacing the appropriate 1'*s* in the assignment matrix by -1 (remember that the 1'*s*, and now the -1'*s*, have no numerical value; they are just convenient symbols).

For the example above, this initial assignment gives

$$
\begin{array}{c|cccc|}
I_i \backslash J_j & J_1 & J_2 & J_3 & J_4 \\
\hline
I_1 & -1 & & 1 & \\
I_2 & & -1 & 1 & \\
I_3 & 1 & 1 & & -1 \\
I_4 & 1 & & & \\
\hline
\end{array}
\quad . \qquad (2)
$$

A new capacity function corresponding to this flow would have

$k(s,I_1) = k(s,I_2) = k(s,I_3) = 0$

$k(I_1,J_1) = k(I_2,J_2) = k(I_3,J_4) = K - 1$, which is effectively still K,

$k(J_1,I_1) = k(J_2,I_2) = k(J_4,I_3) = 1$, instead of 0,

$k(J_1,s') = k(J_2,s') = k(J_4,s') = 0$,

and all other capacities unchanged.

In seeking an unsaturated path from s to s', we can go from s to any unassigned individual, I_4 in this case, then to any J_j for which this individual is qualified, then to s' if the job is unassigned or to I_i if J_j has been assigned to I_i, and so on. In terms of the current

matrix **Q** this means we can find an unsaturated path from I_i to any J_j where $q_{ij}=1$ and from J_j to that I_i for which $q_{ij}=-1$; in other words along a row of **Q** to any 1 and then along the column to the -1.

From (2) we find the unsaturated path

$$s \to I_4 \to J_1 \to I_1 \to J_3 \to s'. \tag{3}$$

Adding this flow to the flow with value 3 which led to (2) gives a flow with value 4 and thus all 4 individuals are assigned. It necessitates the following changes to the current **Q**:

$$\begin{aligned} &q_{41} \text{ from } 1 \text{ to } -1,\\ &q_{11} \text{ from } -1 \text{ to } 1,\\ &q_{13} \text{ from } 1 \text{ to } -1,\end{aligned}$$

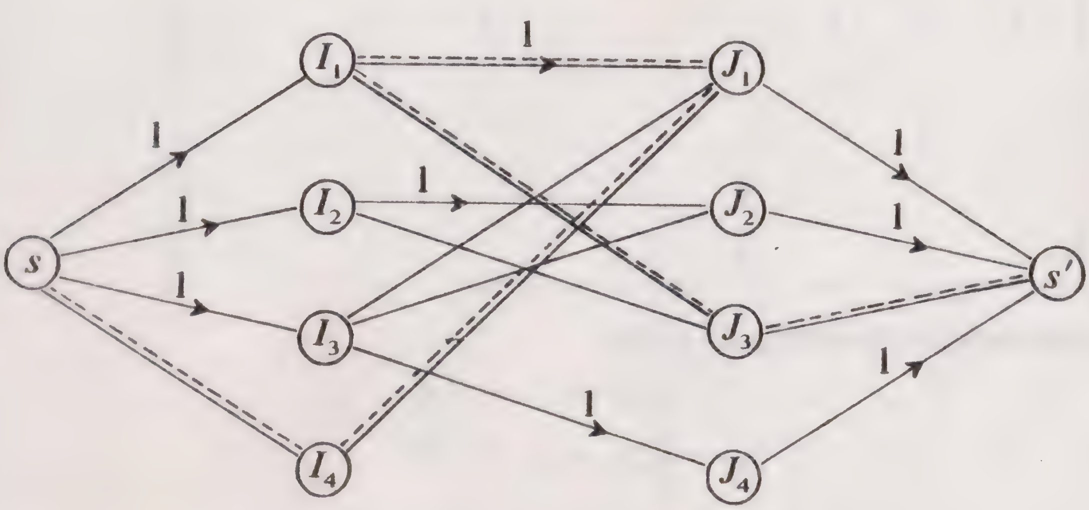
to give

$I_i \backslash J_j$	J_1	J_2'	J_3	J_4
I_1	1		-1	
I_2		-1	1	
I_3	1	1		-1
I_4	-1			

, (4)

which indicates the (final) assignment of all individuals.

The assignment network with the initial flow indicated is



(5)

and the dashed path is that of (3).

A slightly more complicated example has the qualification matrix and initial flow given by

$I_i \backslash J_j$	J_1	J_2	J_3	J_4
I_1	−1	1		1
I_2		−1	1	
I_3	1			−1
I_4	1	1		

, (6)

and the unsaturated path search given by

$$s \rightarrow \{I_4 \rightarrow \begin{cases} J_1 \rightarrow I_1 \rightarrow \{J_4 \rightarrow I_3 \\ J_2 \rightarrow I_2 \rightarrow \{J_3 \rightarrow s'. \end{cases}$$

12.4

A more interesting simple assignment problem is provided by the qualification and initial flow matrix

$I_i \backslash J_j$	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}
I_1				−1	1				1	
I_2								−1	1	1
I_3	−1	1	1			1	1			
I_4									−1	1
I_5				1	−1			1	1	
I_6	1	−1	1				1			
I_7			−1			1	1			
I_8	1	1				−1				
I_9				1	1			1		
I_{10}				1	1				1	

. (1)

The unsaturated path search gives

$$s \rightarrow \begin{cases} I_9 \rightarrow \begin{cases} J_4 \rightarrow I_1 \rightarrow \{X \\ J_5 \rightarrow I_5 \rightarrow \{X \\ J_8 \rightarrow I_2 \rightarrow \{J_{10} \rightarrow s' \end{cases} \\ I_{10} \rightarrow \{J_9 \rightarrow I_4 \end{cases},$$

and the new qualification and flow matrix (2)

$I_i \backslash J_j$	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}
I_1				−1	1				1	
I_2								1	1	−1
I_3	−1	1	1			1	1			
I_4									−1	1
I_5				1	−1			1	1	
I_6	1	−1	1				1			
I_7			−1			1	1			
I_8	1	1				−1				
I_9				1	1			−1		
I_{10}				1	1				1	

. (2)

The unsaturated path search for (2) gives

$$s \to \{I_{10} \to \begin{cases} J_4 \to I_1 \to \{X \\ J_5 \to I_5 \to \{J_8 \to I_9 \to \{X \\ J_9 \to I_4 \to \{J_{10} \to I_2 \to \{X \ . \end{cases} \tag{3}$$

Thus there are no unsaturated paths, a maximum flow has value 9, at most nine individuals can be assigned, and so for some p, $1 \le p \le 10$, there must be a set of p individuals together qualified for $p-1$ $(= p-(10-9))$ jobs.

The minimum cut given by (3) is

$$S = \{s, I_1, I_2, I_4, I_5, I_9, I_{10}, J_4, J_5, J_8, J_9, J_{10}\}$$

and we confirm that the six individuals $I_1, I_2, I_4, I_5, I_9, I_{10}$ are qualified for only the five jobs $J_4, J_5, J_8, J_9, J_{10}$.

12.5 The Optimum Assignment Problem, also known as the Marriage Problem

As we observed in chapter 10, this is a degenerate form of the categorised optimum assignment problem in which each category contains only one individual or job. ~~As with the simple assignment problem, it is convenient for discussion purposes to assume $m = n$, which can easily be arranged. Then the problem may be defined as follows~~: As with the transportation problem where we ensured that $\sum_i b_i = \sum_j d_j$ by introducing a fictitious source or destination with zero cost coefficients, here we can easily arrange that $m=n$ by introducing fictitious persons or jobs as necessary with zero ratings. It is convenient for discussion purposes to assume that this has already been done so that $m=n$. Then the problem may be defined as follows:

> n individuals $I_1, I_2, \ldots, I_n$ apply for n jobs $J_1, J_2, \ldots, J_n$ and the i-th individual's rating for the j-th job is r_{ij}; find the assignment scheme which maximises the sum of the assigned ratings.

This is the same as finding n elements of the $n \times n$ rating matrix **R** such that exactly one is in each row and column of **R** and their sum is maximised, which is itself the same as finding that permutation matrix **P** (see section 3.7) which maximises the trace of **PR** (or of **RP**), (trace $(\mathbf{A}) = \Sigma_{i=1}^{n} a_{ii}$).

Both are the same as the *l.p.p.*

$$\textit{maximise} \quad \Sigma_{i,j=1}^{n} r_{ij} x_{ij} \quad \textit{subject to} \tag{1}$$
$$\Sigma_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, n; \ \Sigma_{i=1}^{n} x_{ij} = 1, \quad j = 1, 2, \ldots, n,$$
$$\textit{and } x_{ij} \geq 0,\ i,j = 1, 2, \ldots, n.$$

The method of chapter 10 for solving transportation problems can clearly be used and will result in an integer solution since d_i, $i = 1, 2, \ldots, m$, and b_j, $j = 1, 2, \ldots, n$, of (3) section 10.2 here all have the value 1. However, all *b.f.s.s* will have exactly n non-zero basic variables and exactly n zero basic variables, and although the procedure of section 10.4(iii) to avoid cycling can be used without difficulty, we are still likely to obtain $\theta = 0$ frequently, and each such stage produces no definite progress towards the optimum assignment. For the case in which the ratings are integers an interesting alternative method, which combines the duality theorem with the method of section 11.3 for the simple assignment problem, is developed in this and the following section. Before this development begins we mention a piquant interpretation of the optimum assignment problem, *the marriage problem*. Here, a community of n men and n women, members of a pioneering colony perhaps, decide that the future happiness of the community (and, no doubt, its present tranquility) would best be assured by abandoning the traditional haphazard and competitive process of courtship in favour of an orderly and fair assignment. Accordingly, each woman expresses the desirability of each of the men as a marriage partner by choosing for each of them a numerical rating (integer) and it is agreed that the assignment which maximises the sum of the assigned ratings will provide the basis for a comprehensive ceremony. As the example in section 12.7 demonstrates, "total antipathy" can easily be taken into account.

Instead of converting the *l.p.p.* (1) to canonical form and using results from chapters 5 and 10 we shall give an independent proof of the duality theorem, which leads directly to a computational algorithm. In this instance, we will treat the problem directly as a

maximisation problem so that the ratings can be left as positive integers and a minus sign used to denote an assignment as in section 12.3.

The *l.p.p.* (1) may be written

$$\textit{maximise} \quad \mathbf{r}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{Ax} = \mathbf{e},\ \mathbf{x} \geq \mathbf{0}, \tag{2}$$

where $\mathbf{A}$ is the $2n \times n^2$ transportation matrix and $\mathbf{e}$ is a $2n$-vector with $e_i = 1$, $i = 1, 2, \ldots, 2n$. We will refer to the *l.p.p.* (2) as the primal; its dual is

$$\textit{minimise} \quad \Sigma_{i=1}^n u_i + \Sigma_{j=1}^n v_j \quad \textit{subject to} \quad u_i + v_j \geq r_{ij},$$
$$i, j = 1, 2, \ldots, n\ (ER). \tag{3}$$

The objective functions of both problems are bounded above and below respectively (*ER*) and both problems have feasible solutions, so both problems have optimum solutions. The duality theorem is therefore simplified, but has the additional complication of integer requirements. We shall state it in terms of the *l.p.p.s* (2) and (3).

12.6 Theorem 15. The Duality Theorem for the Integer Optimum Assignment Problem

The maximum value of $f(\mathbf{x}) = \Sigma_{i,j=1}^n r_{ij}x_{ij}$, where the r_{ij} are integers, subject to

$$\Sigma_{i=1}^n x_{ij} = 1,\ j = 1, 2, \ldots, n,\ \ \Sigma_{j=1}^n x_{ij} = 1,\ i = 1, 2, \ldots, n, \text{ and}$$
$$x_{ij} = 0 \text{ or } 1, \quad i, j = 1, 2, \ldots, n,$$

is the same as the minimum value of $g(\mathbf{u},\mathbf{v}) = \Sigma_{i=1}^n u_i + \Sigma_{j=1}^n v_j$ subject to $u_i + v_j \geq r_{ij}$ and u_i, v_j integers, $i, j = 1, 2, \ldots, n$ ■

The converse result, that $f(\mathbf{x}) = g(\mathbf{u},\mathbf{v})$ for feasible $\mathbf{x}$, $\mathbf{u}$, $\mathbf{v}$ implies optimality, is easily established directly. For *any* feasible $\mathbf{x}$, $\mathbf{u}$, $\mathbf{v}$

$$\Sigma_{i,j} r_{ij}x_{ij} \leq \Sigma_{i,j}(u_i + v_j)x_{ij} = \Sigma_{i,j} u_i x_{ij} + \Sigma_{i,j} v_j x_{ij}$$
$$= \Sigma_i u_i \Sigma_j x_{ij} + \Sigma_j v_j \Sigma_i x_{ij} = \Sigma_i u_i + \Sigma_j v_j,$$

so that $\max f(\mathbf{x}) \leq \min g(\mathbf{u},\mathbf{v})$.

To establish the theorem itself, suppose $\mathbf{u}$ and $\mathbf{v}$ satisfy the constraints of the *l.p.p.* (3) and define a qualification matrix $\mathbf{Q}$ by

$$q_{ij} = \begin{cases} 1 & \text{if} \quad u_i + v_j = r_{ij}, \\ 0 & \text{if} \quad u_i + v_j > r_{ij},\ i, j = 1, 2, \ldots, n. \end{cases}$$

For the simple assignment problem defined by $\mathbf{Q}$ either (i) all I_i can be assigned, or (ii) not all I_i can be assigned, and we examine these two possibilities in turn.

(i) Perform the assignment and put $x_{ij} = 1$ if I_i is assigned to J_j, and $x_{ij} = 0$ otherwise, $i, j = 1, 2, \ldots, n$.

Then

$$\Sigma_{i,j} r_{ij} x_{ij} = \Sigma_{\substack{i,j \text{ such} \\ \text{that } x_{ij} \neq 0}} r_{ij} = \Sigma_{\substack{i,j \text{ such} \\ \text{that } x_{ij} \neq 0}} (u_i + v_j) = \Sigma_{i=1}^{n} u_i + \Sigma_{j=1}^{n} v_j,$$

because for each j there is exactly one $x_{ij} \neq 0$, and for each i there is exactly one $x_{ij} \neq 0$, and hence the assignment is optimum.

(ii) In this case, we know from theorem 14 that there must be a subset, P say, of p of the individuals I_i who are together qualified for a subset, T say, of t of the jobs J_j, where $t < p$.

Define new values $\mathbf{u}'$, $\mathbf{v}'$ for the dual variables by

$$u_i' = \begin{cases} u_i - 1 & \text{if} \quad I_i \in P \\ u_i & \text{if} \quad I_i \notin P, \end{cases}$$

$$v_j' = \begin{cases} v_j + 1 & \text{if} \quad J_j \in T \\ v_j & \text{if} \quad J_j \notin T. \end{cases}$$

The new values of the dual variables satisfy the dual constraints $u_i' + v_j' \geq r_{ij}$, because

if $I_i \in P$ and $J_j \in T$ then

$$u_i' + v_j' = u_i - 1 + v_j + 1 = u_i + v_j \geq r_{ij},$$

if $I_i \notin P$ and $J_j \notin T$ then

$$u_i' + v_j' = u_i + v_j \geq r_{ij},$$

if $I_i \notin P$ and $J_j \in T$ then

$$u_i' + v_j' = u_i + v_j + 1 > u_i + v_j \geq r_{ij}.$$

If $I_i \in P$ and $J_j \notin T$ then I_i is not qualified for J_j, so $q_{ij} = 0$ and $u_i + v_j > r_{ij}$. Since u_i, v_j, r_{ij} are integers $u_i + v_j - 1 \geq r_{ij}$, and $u_i' + v_j' = u_i - 1 + v_j \geq r_{ij}$.

Thus $u_i' + v_j' = u_i - 1 + v_j \geq r_{ij}$ for $i, j = 1, 2, \ldots, n$. However

$$\Sigma_i u_i' + \Sigma_j v_j' = (\Sigma_i u_i) - p + (\Sigma_j v_j) + t < \Sigma_i u_i + \Sigma_j v_j,$$

because $p > t$, and this contradicts the optimality of u, $\mathbf{v}$ ∎

The proof of theorem 15 provides a method for solving the marriage problem. An initial feasible solution for the dual is easy to find, for example

$$v_j = 0, \quad u_i = \max_{j=1,\ldots,n} r_{ij}.$$

Then if all individuals in the simple assignment problem corresponding to this dual solution can be assigned, the assignment solves the marriage problem, and if not the sets P and T lead to the improved dual solution $\mathbf{u}'$, $\mathbf{v}'$.

12.7 Example

Solve the marriage problem with $m = n = 5$ and rating matrix

$$\begin{pmatrix} 12 & 9 & 10 & 3 & 8 \\ 6 & 6 & 2 & z & 9 \\ 6 & 8 & 10 & 11 & 9 \\ 6 & 3 & 4 & 1 & z \\ 11 & 1 & 10 & 9 & 12 \end{pmatrix}, \tag{1}$$

where z denotes antipathy, i.e. an unacceptable assignment.

An initial dual solution is given by $v_j = 0$, $j = 1, 2, \dots, 5$, $u_1 = 12$, $u_2 = 9$, $u_3 = 11$, $u_4 = 6$, $u_5 = 12$, and we solve the simple assignment problem with qualification matrix $\mathbf{Q}$,

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2}$$

We can see at once that all 5 individuals are together qualified for only J_1, J_4, J_5 so we can decrease u_1, u_2, u_3, u_4, u_5 and increase v_1, v_4, v_5. This is the simplest way to improve the dual solution when not all jobs have a qualified applicant.

Denoting those r_{ij} for which $u_i + v_j = r_{ij}$ by $^*r_{ij}$ we now have

$I_i \backslash J_j$	J_1	J_2	J_3	J_4	J_5	
I_1	*12	9	10	3	8	11 ↓
I_2	6	6	2	z	*9	8 ↓
I_3	6	8	*10	*11	9	10 ↓
I_4	*6	3	4	1	z	5 ↓
I_5	11	1	10	9	*12	11 ↓
	1↑	0	0↑	1↑	1↑	$v_j \backslash u_i$

, (3)

and again, J_2 has no qualified individuals so the u_i and v_j indicated by ↓ and ↑ are decreased and increased respectively. In this case a decrease/increase of 1 produces no new $^*r_{ij}$ so we can do the next stage as well if we decrease/increase by 2.

This gives

$I_i \backslash J_j$	J_1	J_2	J_3	J_4	J_5	
I_1	$-$*12	*9	10	3	8	9↓
I_2	6	$-$*6	2	z	*9	6↓
I_3	6	*8	$-$*10	*11	9	8
I_4	*6	*3	4	1	z	3↓
I_5	11	1	10	9	$-$*12	9↓
	3↑	0↑	2	3	3↑	$v_j \backslash u_i$

(4)

Now all J_j have a qualified individual, so to solve the simple assignment problem we assign initially I_1 to J_1, I_2 to J_2, I_3 to J_3, I_5 to J_5 and search for an unsaturated path from s to I_4 and hence to s'. This initial flow has been indicated in (4) by inserting a minus sign before ratings r_{11}, r_{22}, r_{33}, r_{55}.

$$s \to \{I_4 \to \begin{cases} J_1 \to I_1 \to \{X \\ J_2 \to I_2 \to \{J_5 \to I_5 \to \{X. \end{cases}$$

Thus $P = \{I_1, I_2, I_4, I_5\}$ and $T = \{J_1, J_2, J_5\}$, so we decrease u_1, u_2, u_4, u_5, and increase v_1, v_2, v_5 as indicated in (4).

$I_i \backslash J_j$	J_1	J_2	J_3	J_4	J_5	
I_1	*$-$12	*9	*10	3	8	8
I_2	6	$-$*6	2	z	*9	5
I_3	6	8	$-$*10	*11	9	8
I_4	*6	*3	*4	1	z	2
I_5	11	1	*10	9	$-$*12	8
	4	1	2	3	4	$v_j \backslash u_i$

(5)

Notice that in (4) $u_3 + v_2 = r_{32}$, but in (5) $u_3 + v_2 > r_{32}$, so that qualifications are not necessarily maintained from one stage to the next.

The initial assignment in (5) is indicated by minus signs, and the unsaturated path search gives

$$s \to \{I_4 \to \begin{cases} J_1 \to I_1 \to \{X \\ J_2 \to I_2 \to \{J_5 \to I_5\,. \\ J_3 \to I_3 \to \{J_4 \to s' \end{cases}$$

This extra flow in the assignment network means that all individuals are assigned,

$$I_1 \text{ to } J_1,\ I_2 \text{ to } J_2,\ I_3 \text{ to } J_4,\ I_4 \text{ to } J_3,\ I_5 \text{ to } J_5. \tag{6}$$

The sum of the assigned ratings is

$$12 + 6 + 11 + 4 + 12 = 45,$$

and the value of the dual solution is

$$(8 + 5 + 8 + 2 + 8) + (4 + 1 + 2 + 3 + 4) = 45,$$

confirming that the assignment (6) is optimum.

12.8

We should not leave the problems of this and the previous chapter without some further comments. The assignment problems we have discussed are examples of *combinatorial* or *set-covering* problems (sometimes called *zero-one* problems). For such problems a variety of methods has been devised, each of which is more or less efficient depending on the structure of the particular problem in question (see, for example, {9}, {10}, {11}). In chapters 11 and 12 we have presented methods for network flow, simple assignment and optimum assignment problems partly for the intrinsic appeal of the problems themselves, and partly for the interesting way in which the methods develop from each other.

A common feature both of the methods developed here and of methods for integer *l.p.p.s* in general is the need for repeated extensive searches through stored information, and these searches can be very time-consuming. This feature distinguishes the situation from that of solving general *l.p.p.s* by the simplex method. The optimism that has been expressed earlier about the efficiency of the simplex method in practice cannot always be carried over to these more specialised problems; in some instances the exponential-time quality of the algorithms is experienced in practice. There is not a contradiction here because the restriction to integer values is qualitatively rather different from the general half-space constraints. Notice also that even small network or assignment problems have many constraints and variables and give rise to quite large *l.p.p.s* (see exercise 11.1).

Exercises 12

1. Solve the simple assignment problem whose qualification matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

   (i) by inspection (constructively),
   (ii) by the method of section 12.3.

2. Solve the simple assignment problem whose qualification matrix is given by

$I_i \backslash J_j$	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}
I_1				1		1		1		
I_2	1		1		1					1
I_3					1		1		1	1
I_4	1					1				
I_5	1	1				1				
I_6		1		1				1		
I_7			1				1		1	
I_8		1		1		1		1		
I_9			1		1		1		1	1
I_{10}				1		1		1		

3. (i) The size and shape of the *tree* that describes the search for an unsaturated path is not known before the search takes place. Discuss how the search procedure could be implemented and stored automatically by a computer.
   (ii) From the point of view of automatic computation, why is it advantageous to denote assignments (flows in the assignment network) by minus signs, and how could the *equality dual constraints* $^*r_{ij}$ be denoted?

4. Solve the optimum assignment problem in which the rating matrix is

$$\begin{pmatrix} 3 & 4 & 2 & 5 & 7 & 8 & 6 \\ 2 & 1 & 0 & 9 & 7 & 2 & 3 \\ 7 & 7 & 6 & 1 & 0 & 0 & 3 \\ 2 & 1 & 5 & 5 & 4 & 7 & 8 \\ 3 & 6 & 5 & 2 & 7 & 8 & 2 \\ 0 & 1 & 0 & 2 & 5 & 7 & 2 \\ 2 & 6 & 6 & 5 & 3 & 2 & 1 \\ 6 & 2 & 5 & 4 & 7 & 7 & 9 \end{pmatrix}.$$

5. Solve the marriage problem in which the rating matrix is

$$\begin{pmatrix} 3 & 7 & 8 & 2 & 1 & z & 5 & 4 & z & 6 \\ 6 & 3 & 5 & 7 & 9 & 2 & 1 & 4 & 2 & 8 \\ 9 & 8 & 3 & 7 & 2 & 4 & 7 & 6 & 1 & 5 \\ 7 & 5 & 6 & 9 & 1 & 3 & 8 & z & 2 & 4 \\ 1 & 10 & 8 & 9 & 5 & z & 6 & 4 & 2 & 7 \\ 4 & 2 & 1 & 6 & z & 5 & 3 & 3 & 7 & 8 \\ 6 & 8 & 10 & 9 & 4 & 3 & 5 & 1 & 7 & 2 \\ 7 & 8 & 4 & 3 & 6 & 2 & 1 & z & 5 & z \\ 3 & 9 & 4 & 2 & 5 & z & 7 & z & 8 & 1 \\ 9 & 3 & 1 & 8 & 4 & z & 7 & 2 & 6 & 5 \end{pmatrix},$$

where z denotes total antipathy.

6. The marriage problem was described from a female point of view in section 12.5. Suppose the men involved wish their opinions (ratings) to be taken into account as well (not instead). Suggest two distinct ways in which this could be done. (One way is perhaps rather unrealistic, but is more realistic for the applicants and jobs situation.)
7. In the optimum assignment problem prove that in an optimum assignment at least one individual is assigned to the job he or she is best qualified (highest rated) for. Deduce another similar result.

   Hint: assume a convenient form for the assignment, consider the cases $n = 2, 3, 4$ and obtain the general result by induction.
8. Devise an improvement to the method of section 12.3 for obtaining an initial assignment for solving the simple assignment problem.

NOTES

CHAPTER 13

GAME THEORY: TWO-PERSON MATRIX GAMES

13.1

For a certain class of games the problem of determining the best stratagem for playing the game can be formulated as a *l.p.p.* These are *two-person zero-sum matrix games.*

A simple example enjoyed by children is the stone-paper-scissors game. Here, the two players X and Y simultaneously shout one of the words stone, paper, or scissors; if both shouts are the same, the result is a tie, otherwise stone beats scissors, scissors beats paper and paper beats stone. The game is played many times and the winner each time receives a fixed predetermined "reward". In general the essential features of a two-person zero-sum matrix game are

(i) the two players compete against each other with no external influences,
(ii) each play of the game consists of both players choosing independently one of a finite number of possible alternatives,
(iii) the consequence of any pair of choices is fixed and known in advance, and
(iv) the winner's gain is the loser's loss.

Such a game is completely described by an $m \times n$ matrix $\mathbf{A}$, the *payoff matrix,* in which the element a_{ij} is the *payoff,* i.e. X's gain and Y's loss, when X chooses the j-th of X's n possible alternatives and Y chooses the i-th of Y's m possible alternatives.

For the stone-paper-scissors game the payoff matrix is:

$$\begin{array}{ccc|c|c|c|l}
 & & & \multicolumn{3}{c}{X \textit{ plays}} & \\
 & & & \textit{stone} & \textit{paper} & \textit{scissors} & \\
\cline{4-6}
 & \textit{stone} & & 0 & 1 & -1 & \\
\cline{4-6}
Y \textit{ plays} & \textit{paper} & & -1 & 0 & 1 & = \mathbf{A}, \qquad (1) \\
\cline{4-6}
 & \textit{scissors} & & 1 & -1 & 0 & \\
\cline{4-6}
\end{array}$$

where the loser each time pays the winner one point. If we regard $\mathbf{A}$ as defining the game from X's point of view, then $-\mathbf{A}$ defines

the game from *Y's* point of view. Thus a game defined by a skew-symmetric payoff matrix, $\mathbf{A}^T = -\mathbf{A}$, is the same for both players and we would expect such a game to be fair (see exercise 13.2).

The definitive properties of a two-person zero-sum matrix game probably apply precisely only in genuine games, but by categorising the possible alternative actions and assessing the payoffs, a number of economic, management and military situations can be modelled and analysed as matrix games (e.g. see exercise 13.8).

The results concerning matrix games were first established by von Neumann in 1928, long before *l.p.* theory was developed. They are easily obtained when the connection with *l.p.* has been made and we shall confine ourselves to this approach, which was first established by Dantzig in 1951.

If either player can predict the other's next choice then, because the payoffs are known and each play of the game is a separate independent event, that player will use that information to his or her advantage. For example, in the stone-paper-scissors game, if X knows that *Y's* next play will be stone then X will play paper. For this reason, both players must make each successive choice randomly. However, within this restriction they can both decide the proportion of times they choose each of their possible alternatives. So if x_j, $j = 1, 2, \ldots, n$, is the probability that X plays *X's* j-th alternative, then *X's* problem is to choose that strategy vector $\mathbf{x}$ such that *X's* average gain is a maximum; and *Y's* problem is to choose a strategy vector $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$ such that *Y's* average loss is a minimum.

If X chose the strategy vector $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})^T$, for the stone-paper-scissors game, then X would play:

stone, on average once every two plays,

paper, on average once every four plays, and

scissors, on average once every four plays.

This is not *X's* optimum stratagem, because if Y played paper every time then $\mathbf{y} = (0, 1, 0)^T$ and *X's* average gain would be $-\frac{1}{4}$, whereas if X chose $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ (which *is* *X's* optimum strategem) then whatever Y played, X would expect to win, lose or draw equally often, so *X's* average gain would be zero. These examples of stratagems make it clear that we need to be precise about what we mean by optimum stratagems for X and Y. We said that $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ was optimum for X because with this stratagem X can expect to break even. With any stratagem for which some $x_i > \frac{1}{3}$, Y can choose a stratagem such that X can expect, on average, to lose: for example $\mathbf{y} = (0, 1, 0)^T$

if $x_1 > \frac{1}{3}$ and $x_2, x_3 < \frac{1}{3}$. However, if Y chooses $(0,1,0)^T$ then $\mathbf{x} = (0,0,1)^T$ ensures that X wins 1 every time. So, by an *optimum stratagem for X*, we mean that stratagem which maximises X's average gain given that Y will choose a stratagem which minimises Y's average loss. Thus we assume that both players will play consistently and as skilfully as possible, and both will choose stratagems according to this assumption. Their choice is based on an assumption about the nature of their opponent's stratagem, but is independent of it and made without observing it. This rather subtle notion distinguishes two-person matrix games from similar problems in which the payoff matrix refers to a person-nature or a person-machine situation.

The idea of optimum stratagems as we have defined it is only compatible with the zero-sum aspect of matrix games if there is a unique quantity v, which we call the *value of the game*, and which is both the maximum amount X can be sure of gaining on average and the minimum amount Y cannot avoid losing on average. The existence of such a v and of optimum stratagems for X and Y is the substance of the *Fundamental Theorem of Two-person Zero-sum Matrix Games.* (Such games from now on will simply be called *matrix games.*) A *fair* game is one which has value zero.

13.2 The Linear Programming Connection

Consider the problem of determining X's optimum stratagem. The variables x_j, $j = 1, 2, \ldots, n$, as they are probabilities (or proportions), must satisfy

$$\Sigma_{j=1}^{n} x_j = 1, \;\; x_j \geq 0, \;\; j = 1, 2, \ldots, n. \tag{1}$$

For those plays in which Y chooses Y's i-th alternative, X's expectation, or average gain, is

$$\Sigma_{j=1}^{n} a_{ij} x_j .$$

Denoting by p the average gain which the stratagem $\mathbf{x}$ will guarantee to X, we must have

$$\Sigma_{j=1}^{n} a_{ij} x_j \geq p, \;\; i = 1, 2, \ldots, m. \tag{2}$$

Thus X's problem is to choose $x_1, x_2, \ldots, x_n, p$ so as to *maximise* $p = (0, 0, \ldots, 0, 1)\begin{pmatrix} \mathbf{x} \\ p \end{pmatrix}$ subject to the constraints (1) and (2). This is the *l.p.p.*

$$\textit{maximise } (\mathbf{0}^T, 1)\begin{pmatrix} \mathbf{x} \\ p \end{pmatrix} \textit{ subject to } \mathbf{A}\mathbf{x} \geq p\mathbf{e}, \;\; \mathbf{e}^T\mathbf{x} = 1, \;\; \mathbf{x} \geq \mathbf{0}. \tag{3}$$

The same argument applied to Y's problem (ER) gives the *l.p.p.*

$$\Sigma_{i=1}^{m} y_i = 1,\ y_i \geq 0,\ i = 1, 2, \ldots, m, \tag{4}$$
$$\Sigma_{i=1}^{m} y_i a_{ij} \leq q,\ j = 1, 2, \ldots, n,\ \textit{minimise } q,$$

i.e. *minimise* $(\mathbf{y}^T, q)\begin{pmatrix}\mathbf{0}\\1\end{pmatrix}$ *subject to* $\mathbf{y}^T\mathbf{A} \leq q\mathbf{e}^T, \mathbf{y}^T\mathbf{e} = 1, \mathbf{y} \geq \mathbf{0}$, (5)

where q is Y*'s* expected loss, and $-q$ is Y*'s* expected gain.

The *l.p.p.* (5) is the dual of the *l.p.p.* (3) (*ER*), both have feasible solutions, trivially, and so both have optimum solutions, with *maximum* $p =$ *minimum* $q = v$ say. Thus the fundamental theorem for matrix games follows at once from the duality theorem for *l.p.p.s*, and the solution of a matrix game (the value of the game and the optimum stratagems) can be obtained by solving a single *l.p.p.* Notice that $v = \mathbf{y}_{opt}^T \mathbf{A}\mathbf{x}_{opt}$ (*ER*).

Instead of solving (3) or (5) by converting to canonical form, we first reformulate both X*'s* problem and Y*'s* problem.

The optimum stratagems, $\mathbf{x}_0$ and $\mathbf{y}_0$ say, are unchanged if we replace a_{ij} by $a_{ij} + \alpha$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, but the value of the game changes, by an increase α (*ER*). If we choose α to ensure that all $a_{ij} + \alpha$ are strictly positive then the value of the game must be strictly positive, and so in (3) and (5) p and q are positive. Assuming that this has been done and denoting $a_{ij} + \alpha$ by a'_{ij}, put $x'_j = x_j/p$, $j = 1, 2, \ldots, n$, so that X*'s* problem becomes

$$\textit{minimise}\quad \mathbf{e}^T\mathbf{x}'\quad \textit{subject to}\quad \mathbf{A}'\mathbf{x}' \geq \mathbf{e},\ \mathbf{x}' \geq \mathbf{0}, \tag{6}$$

because $\mathbf{e}^T\mathbf{x}' = \Sigma_{j=1}^{n} x_j/p = 1/p$, and X wishes to maximise p. This is a *l.p.p.* in standard primal form.

Similarly, with $y'_i = y_i/q$, Y*'s* problem becomes

$$\textit{maximise}\quad \mathbf{y}'^T\mathbf{e}\quad \textit{subject to}\quad \mathbf{y}'^T\mathbf{A}' \leq \mathbf{e}^T,\ \mathbf{y} \geq \mathbf{0}, \tag{7}$$

which is in standard dual form.

In this formulation the value of the game v is given by $(\Sigma_{j=1}^{n} x'_j)^{-1}$ or $(\Sigma_{i=1}^{n} y'_i)^{-1}$ and the optimum stratagems by $\mathbf{x}_0 = v\mathbf{x}'_0$, $\mathbf{y}_0 = v\mathbf{y}'_0$.

Example

We *verify* for the stone-paper-scissors game that both the optimum stratagems are $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ and the value of the game is 0.

With

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

and $\mathbf{x}_0 = \mathbf{y}_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$, we have

$$\mathbf{A}\mathbf{x}_0 = \mathbf{0},\ \mathbf{y}_0^T\mathbf{A} = \mathbf{0}^T,\ \text{i.e.}\quad \mathbf{A}\mathbf{x}_0 \geq 0\mathbf{e},\ \mathbf{y}_0^T\mathbf{A} \leq 0\mathbf{e}^T,$$
$$\mathbf{x}_0 \geq \mathbf{0},\ \mathbf{y}_0 \geq \mathbf{0},\ \mathbf{e}^T\mathbf{x}_0 = \mathbf{y}_0^T\mathbf{e} = 1,\ p = q = 0.$$

Thus $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$ is a feasible solution for both (3) and (5); it gives the same value for both objective functions and is therefore the optimum solution for both problems.

Alternatively, choosing $\alpha = 2$, we have

$$\mathbf{A}' = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}. \tag{8}$$

With $\mathbf{x}_0' = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6})^T = \mathbf{y}_0'$, we have

$$\mathbf{A}'\mathbf{x}_0' = \mathbf{e},\ \mathbf{y}_0'^T\mathbf{A}' = \mathbf{e}^T,\ \mathbf{e}^T\mathbf{x}_0' = \tfrac{1}{2},\ \mathbf{y}_0'^T\mathbf{e} = \tfrac{1}{2},\ \text{so}$$

$v = 2,\ \mathbf{x}_0 = \mathbf{y}_0 = 2(\frac{1}{6}, \frac{1}{6}, \frac{1}{6})^T = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$, and the value of the game is $2 - \alpha = 0$.

13.3 Pure, Mixed, Dominated and Essential Stratagems; Saddle Point Games

A stratagem $\mathbf{x}$ of the form $\mathbf{x} = \mathbf{e}_j$, $i \leq j \leq n$, is called a *pure stratagem* and means that X chooses the same alternative every time; otherwise a stratagem in which X uses more than one alternative is called a *mixed stratagem.*

A game in which the optimum stratagems for X and Y are pure stratagems is called a *saddle point game* and is easy to recognise.

The optimum pure stratagem for X is the j_0-th, where

$$\max_j (\min_i a_{ij}) \quad \text{is attained with} \quad j = j_0.$$

Similarly the optimum pure stratagem for Y is the i_0-*th,* where

$$\min_i (\max_j a_{ij}) \quad \text{is attained with} \quad i = i_0.$$

If $\max_j(\min_i a_{ij}) = a_{i_0 j_0} = \min_i(\max_j a_{ij})$ then $\mathbf{e}_{j_0}$ and $\mathbf{e}_{i_0}$ are X'*s* and Y'*s* optimal stratagems because they are feasible solutions for (3) and (5) of section (2) with $p = q = a_{i_0 j_0}$.

This observation leads to the alternative formulation of X'*s* problem and Y'*s* problem. For any chosen stratagem $\mathbf{x}$, X'*s* expectations for Y'*s* various alternative plays are given by the vector $\mathbf{A}\mathbf{x}$, and so if Y chooses the stratagem $\mathbf{y}$, X'*s* average gain is

$$\mathbf{y}^T\mathbf{A}\mathbf{x}. \tag{1}$$

The stratagem $\mathbf{y}$ that Y chooses will be that which minimises X'*s* average gain, so that subject to the constraints (1) and (4) of section

13.2, X requires $\mathbf{x}_0$ such that

$$\max_{\mathbf{x}} (\min_{\mathbf{y}} \mathbf{y}^T\mathbf{A}\mathbf{x}) = \min_{\mathbf{y}} \mathbf{y}^T\mathbf{A}\mathbf{x}_0. \tag{2}$$

Similarly, Y requires $\mathbf{y}_0$ such that

$$\min_{\mathbf{y}} (\max_{\mathbf{x}} \mathbf{y}^T\mathbf{A}\mathbf{x}) = \max_{\mathbf{x}} \mathbf{y}_0^T\mathbf{A}\mathbf{x}. \tag{3}$$

The assertion that the quantities

$$\max_{\mathbf{x}} (\min_{\mathbf{y}} \mathbf{y}^T\mathbf{A}\mathbf{x}) \quad \text{and} \quad \min_{\mathbf{y}} (\max_{\mathbf{x}} \mathbf{y}^T\mathbf{A}\mathbf{x})$$

both exist and have the same value, v, is the *minimax theorem* of von Neumann. We have already established this result in the previous section as the fundamental theorem of matrix games, and we know that

$$v = \mathbf{y}_0^T\mathbf{A}\mathbf{x}_0.$$

We observe that

$$\begin{aligned}\min_{\mathbf{y}} (\max_{\mathbf{x}} \mathbf{y}^T\mathbf{A}\mathbf{x}) = \max_{\mathbf{x}} \mathbf{y}_0^T\mathbf{A}\mathbf{x} &\le \mathbf{y}_0^T\mathbf{A}\mathbf{x}_0 \le \min_{\mathbf{y}} \mathbf{y}^T\mathbf{A}\mathbf{x}_0 \\ &= \min_{\mathbf{y}} (\max_{\mathbf{x}} \mathbf{y}^T\mathbf{A}\mathbf{x}),\end{aligned}$$

and that $\mathbf{y}_0^T\mathbf{A}\mathbf{x} \le \mathbf{y}_0^T\mathbf{A}\mathbf{x}_0 \le \mathbf{y}^T\mathbf{A}\mathbf{x}_0$.

Suppose that

$$a_{ij} \ge a_{ik}, \quad i = 1, 2, \ldots, m,$$

then whichever play Y chooses, X never gains more by choosing *X's* k-th alternative play in preference to the j-th. In this situation, we can be sure that $(\mathbf{x}_0)_k = 0$ and we may reduce the size of the payoff matrix $\mathbf{A}$ by removing the k-th column. We say that *X's* k-th alternative, or k-th pure stratagem, is *dominated* by the j-th. Similarly, *Y's* k-th pure strategem is dominated by the i-th if

$$a_{ij} \le a_{kj}, \quad j = 1, 2, \ldots, n,$$

in which case $(\mathbf{y}_0)_k = 0$ and the k-th row of $\mathbf{A}$ may be removed.

It may be the case that after removal of dominated stratagems the reduced payoff matrix reveals dominated stratagems which were not apparent in the original payoff matrix, and significant simplification of a matrix game may result from the elimination of dominated stratagems.

Example

The *skin-game* devised by Kuhn has payoff matrix

$$\begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & -1 \\ -2 & 1 & 0 \end{pmatrix}$$

and appears at first glance to be fair. However, X's first pure stratagem is dominated by the third which reduces the payoff matrix to

$$\begin{pmatrix} -1 & 2 \\ 1 & -1 \\ 1 & 0 \end{pmatrix},$$

and now Y's third stratagem is dominated by the second. The payoff matrix is now

$$\begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix},$$

which indicates a bias towards X.

An *essential* alternative, or *essential* pure stratagem, is one which is used a strictly positive proportion of times in an optimum stratagem. It is not correct that all pure stratagems are either essential or dominated, since a pure stratagem can be dominated by a combination of other alternatives, but not dominated by any one of them (see exercise 13.7).

Extensions to n-person games may be found in {10}.

Exercises 13

1. Prove that the optimum stratagems for a matrix game with payoff matrix **A** are unchanged if **A** is replaced by $\alpha\mathbf{A}$ for some $\alpha > 0$. What is the practical interpretation of this result?
2. Prove that the value of a matrix game in which the payoff matrix is skew-symmetric, $\mathbf{A}^T = -\mathbf{A}$, has value zero. What can you say about the optimum stratagems for X and Y?
3. Verify that the solution of the skin-game (see section 13.3) is $(\frac{1}{5}, (0, \frac{3}{5}, \frac{2}{5})^T, (\frac{2}{5}, \frac{3}{5}, 0)^T)$.
4. Solve the matrix game with payoff matrix
$$\mathbf{A} = \begin{pmatrix} 2 & 1 & -2 \\ -1 & 0 & -1 \\ -2 & 1 & 2 \end{pmatrix}.$$
5. In the matching pennies game the two players X and Y simultaneously uncover a penny: if both coins show heads or both tails X wins and takes both, and if they show one head and one tail Y wins and takes both. Solve this game.

   Suppose the payoff matrix is changed to
$$\begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}$$
   and Y agrees to play only if X pays Y a premium of 1 every 10 plays. Should X agree?
6. *Verify* that the value of the matrix game with payoff matrix
$$\begin{pmatrix} -1 & 2 & 3 \\ 1 & -2 & 3 \\ 1 & 2 & -3 \end{pmatrix}$$
   is $\frac{6}{11}$ and the optimum stratagems are
$$(\tfrac{6}{11}, \tfrac{3}{11}, \tfrac{2}{11})^T \quad \text{and} \quad (\tfrac{5}{22}, \tfrac{8}{22}, \tfrac{9}{22})^T.$$
7. Devise a matrix game in which $m = 2$, $n = 3$, none of X's pure stratagems is dominated, but only two are essential.
8. The hide-and-seek game: One player, Y say, can hide in any element b_{ij} of an $s \times t$ matrix **B**; X chooses to search either a row of **B** or a column of **B**. If X finds Y the payoff is b_{ij}, otherwise the payoff is $-\alpha$. Describe the payoff matrix which defines this situation as a matrix game and discuss situations which can be modelled by this game.

NOTES

NOTES

CHAPTER 14

FURTHER APPLICATIONS: QUADRATIC PROGRAMMING; FUNCTIONAL APPROXIMATION; MATRIX EIGENVALUE PERTURBATION ANALYSIS

14.1

Applications of linear programming techniques discussed in earlier chapters concern situations in which a mathematical description of the problem leads naturally to a *l.p.p.* This chapter is concerned with several problems for which one would not immediately expect linear programming to be useful. The characteristic feature of the problems is the presence of linear constraints, and the methods we develop rely on the fact that the simplex method involves an effective way of handling such information. In each case other methods using different approaches are available which, depending on the particular problem, may be more effective.

Quadratic Programming Problems *(q.p.p.s)*

Here the objective function $f(\mathbf{x})$ which we wish to minimise is a quadratic function of the variables $x_1, x_2, \ldots, x_n$, which we may write

$$f(\mathbf{x}) = \tfrac{1}{2}\,\mathbf{x}^T\mathbf{D}\,\mathbf{x} + \mathbf{c}^T\mathbf{x}, \quad \text{where} \quad \mathbf{D}^T = \mathbf{D} \tag{1}$$

and the problem is to minimise $f(\mathbf{x})$ subject to

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}. \tag{2}$$

A constant term which might be involved in $f(\mathbf{x})$ can be ignored, and any linear inequality constraints may be put in the form (2).

We shall restrict our attention to the case in which $\mathbf{D}$ is positive definite, i.e. $\mathbf{x}^T\mathbf{D}\mathbf{x} > 0$ if $\mathbf{x} \neq \mathbf{0}$. For this case, for any $\mathbf{x} \neq \mathbf{0}$ $f(k\mathbf{x})$ eventually increases without bound as k increases so any *q.p.p.* defined by (1) and (2) has an optimum solution.

In addition to this fact, *q.p.p.s.* differ from *l.p.p.s* by not necessarily attaining their minimum value on the boundary of the feasible region R. To illustrate this, it is convenient (as it was in chapter 1 for *l.p.p.s*) to consider an example in 2-space subject to inequality constraints.

The function $f(\mathbf{x}) = (x_1 - 2)^2 + 2(x_2 - 3)^2$

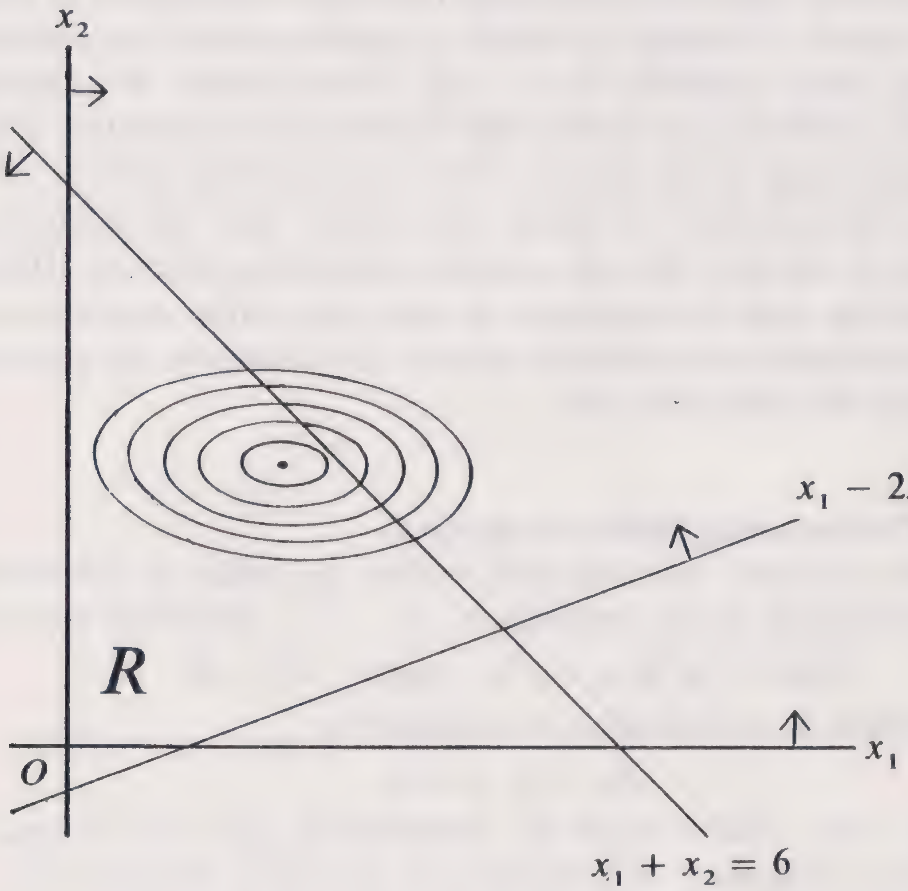
$$= (x_1,x_2)\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} + (-4,-12)\begin{pmatrix}x_1\\x_2\end{pmatrix} + 22,$$

which has no x_1x_2 term, attains its minimum value at the point (2,3) and has a constant value on concentric ellipses centred on this point.

For the set of constraints

$$x_1 + x_2 \leq 6,\ x_1 - 2x_2 \leq 1,\ x_1,x_2 \geq 0$$

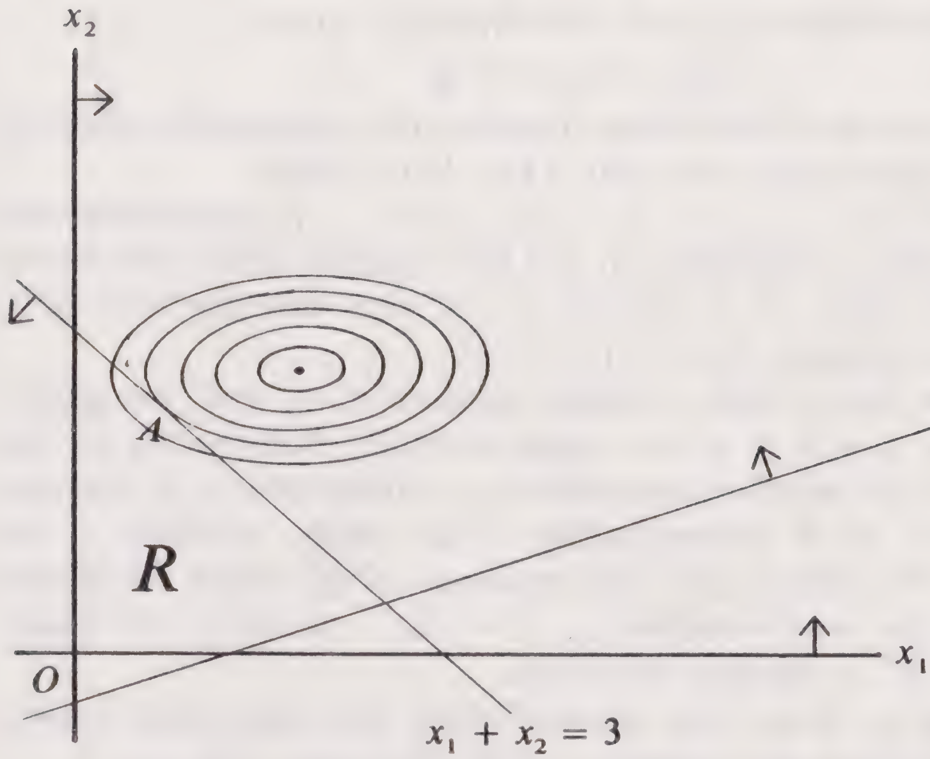
the point (2,3) is an interior point of R, so the constraints are effectively redundant.



For the set of constraints

$$x_1 + x_2 \leq 3,\ x_1 - 2x_2 \leq 1,\ x_1,x_2 \geq 0$$

the point (2,3) is not in R and the minimum value of $f(\mathbf{x})$ subject to these constraints is attained on the boundary of R, at A.



The optimum solution of a *q.p.p.* is characterised by the following result:

Theorem 16

The vector $\mathbf{x}_0$ solves the *q.p.p.* (1), (2) if there are vectors $\boldsymbol{\lambda}_0$ and $\boldsymbol{\mu}_0$ such that $\mathbf{x}_0$, $\boldsymbol{\lambda}_0$, $\boldsymbol{\mu}_0$ satisfy

(i) $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, $\mathbf{x}_0 \geq \mathbf{0}$,

(ii) $\mathbf{D}\mathbf{x}_0 + \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}_0 = \boldsymbol{\mu}_0 \geq \mathbf{0}$,

(iii) $\boldsymbol{\mu}_0^T\mathbf{x}_0 = 0$ ■

To establish this result we suppose that $\mathbf{x}_0$, $\boldsymbol{\lambda}_0$ and $\boldsymbol{\mu}_0$ satisfy (i), (ii), and (iii) and we consider any other feasible vector $\mathbf{x}$.

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}_0) &= \tfrac{1}{2}\,\mathbf{x}^T\mathbf{D}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \tfrac{1}{2}\,\mathbf{x}_0^T\mathbf{D}\mathbf{x}_0 - \mathbf{c}^T\mathbf{x}_0 \\ &= \tfrac{1}{2}\,(\mathbf{x} - \mathbf{x}_0)^T\,\mathbf{D}(\mathbf{x} - \mathbf{x}_0) + \mathbf{c}^T\,(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0^T\mathbf{D}(\mathbf{x} - \mathbf{x}_0). \end{aligned}$$

As $\mathbf{A}\mathbf{x}_0 = \mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{0}$ and $\boldsymbol{\lambda}_0^T\,\mathbf{A}\,(\mathbf{x} - \mathbf{x}_0) = 0$.
Hence

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}_0) &= \tfrac{1}{2}\,(\mathbf{x} - \mathbf{x}_0)^T\,\mathbf{D}(\mathbf{x} - \mathbf{x}_0) + (\mathbf{c}^T + \mathbf{x}_0^T\mathbf{D} + \boldsymbol{\lambda}_0^T\mathbf{A})(\mathbf{x} - \mathbf{x}_0) \\ &= \tfrac{1}{2}\,(\mathbf{x} - \mathbf{x}_0)^T\,\mathbf{D}(\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\mu}_0^T(\mathbf{x} - \mathbf{x}_0). \end{aligned}$$

As $\mathbf{D}$ is positive definite, $\boldsymbol{\mu}_0 \geq \mathbf{0}$, $\mathbf{x} \geq \mathbf{0}$ and $\boldsymbol{\mu}_0^T \mathbf{x}_0 = 0$,

$$f(\mathbf{x}) - f(\mathbf{x}_0) \geq 0 \; \blacksquare$$

This result is just the *Kuhn-Tucker Theorem* for constrained optimisation applied to the *q.p.p.* (see {10}, {12}, {13}, {14}).

If we regard $x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m, \mu_1, \mu_2, \ldots, \mu_n$ as variables we have $(m + n)$ linear constraints (i) and (ii), together with a non-linear constraint (iii) which, as $\mathbf{x} \geq \mathbf{0}$ and $\boldsymbol{\mu} \geq \mathbf{0}$, says that not both of x_j and μ_j may be non-zero, for $j = 1, 2, \ldots, n$.

Suppose we find a basic feasible solution of (i) and we satisfy (iii) by saying $\mu_j = 0$ if x_j is a basic variable. Substituting in the m equations of (ii), which correspond to $\mu_j = 0$ determines $\boldsymbol{\lambda}$, because the m columns of $\mathbf{A}$ corresponding to the basic variables x_j are independent. This leaves $(n - m)$ equations of (ii) which determine the remaining $(n - m)$ variables μ_j, $j = 1, 2, \ldots, n$, for x_j not basic, since $\mathbf{D}$, $\mathbf{X}$, $\mathbf{c}$, $\mathbf{A}^T$, $\boldsymbol{\lambda}$ are now all known.

The solution $\mathbf{x}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ thus obtained does not necessarily satisfy (ii) because we have not ensured $\boldsymbol{\mu} \geq \mathbf{0}$. However, with $\boldsymbol{\mu} = \mathbf{u} - \mathbf{v}$ where $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$, the *l.p.p.*

$$\textit{minimise} \quad v_1 + v_2 + \ldots + v_n \quad \textit{subject to } \text{(i) } \textit{and } \text{(ii)}$$

provides a solution satisfying (i), (ii), and (iii) if $\mathbf{x}^T\mathbf{u} = 0$ and $\mathbf{v}_{opt} = \mathbf{0}$ and thus can be used to solve the *q.p.p.*

14.2

To convert the *l.p.p.* developed in section 14.1 to canonical primal form we have only to put $\boldsymbol{\lambda} = \mathbf{s} - \mathbf{t}$ where $\mathbf{s}, \mathbf{t} \geq \mathbf{0}$.

Assuming that we have already found a *b.f.s.* of $\mathbf{Ax} = \mathbf{b}$ or, more conveniently, that $\mathbf{A} \supset \mathbf{I}_m$ then the *l.p.p.*

$$\textit{minimise} \quad (\mathbf{0}^T, \mathbf{0}^T, \mathbf{0}^T, \mathbf{0}^T, \mathbf{e}^T) \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \\ \mathbf{t} \\ \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

$$\textit{subject to} \quad \mathbf{Ax} = \mathbf{b}, \quad \mathbf{Dx} + \mathbf{A}^T\mathbf{s} - \mathbf{A}^T\mathbf{t} - \mathbf{u} + \mathbf{v} = -\mathbf{c},$$

$$\mathbf{x}, \mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{v} \geq \mathbf{0}$$

can be solved directly by the simplex method, with the modification that, for $j = 1, 2, \ldots, n$, x_j and u_j may not both be basic variables. This ensures that $\mathbf{u}^T\mathbf{x} = 0$ at every stage, and as we will have $\mathbf{v} = \mathbf{0}$

at the optimum stage, $\boldsymbol{\mu}^T\mathbf{x} = (\mathbf{u} - \mathbf{v})^T\mathbf{x} = 0$ at the optimum stage.

The $(m + n) \times (n + m + m + n + n)$ matrix of coefficients is

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{D} & \mathbf{A}^T & -\mathbf{A}^T & -\mathbf{I}_n & \mathbf{I}_n \end{pmatrix}, \text{ where } \mathbf{A} \supset \mathbf{I}_m.$$

This is easily converted to the required form by adding multiples of the first m rows to the last n rows to reduce to zero the columns of $\mathbf{D}$ corresponding to columns of $\mathbf{I}_m$ in $\mathbf{A}$. This does not affect the matrices $-\mathbf{I}_n$ and $\mathbf{I}_n$ in the last n rows, so that the right-hand-side vector, originally $\begin{pmatrix} \mathbf{b} \\ -\mathbf{c} \end{pmatrix}$ now $\begin{pmatrix} \mathbf{b} \\ -\mathbf{c}' \end{pmatrix}$, can be made non-negative by multiplying appropriate rows by -1. Thus initially the $(m + n)$ basic variables will consist of m of $x_1, x_2, \ldots, x_n$ and those n of $u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_n$ which now correspond to columns of $\mathbf{I}_{m+n}$.

Example

$$\begin{aligned} &\textit{Minimise} && x_1^2 + x_2^2 - 8x_1 - 10x_2 \\ &\textit{subject to} && 3x_1 + 2x_2 \le 6,\ x_1, x_2 \ge 0. \end{aligned}$$

Adding a slack variable x_3 to produce an equality constraint

$$3x_1 + 2x_2 + x_3 = 6,\ x_1, x_2, x_3 \ge 0,$$

we have $m = 1$, $n = 3$,

$$\mathbf{A} = (3,2,1),\ \mathbf{D} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ \mathbf{c} = \begin{pmatrix} -8 \\ -10 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\lambda} = \lambda_1,\ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}.$$

An initial *b.f.s.* is $x_1 = x_2 = 0$, $x_3 = 6$, hence $\mu_3 = 0$ and

$$(\boldsymbol{\mu})_3 - (\mathbf{D}\mathbf{x})_3 - (\mathbf{c})_3 = (\mathbf{A}^T\boldsymbol{\lambda})_3,$$

i.e. $0 - 0 - 0 = 1 \times \lambda_1$, so $\lambda_1 = 0$, and then the first and second rows of

$$\mathbf{D}\mathbf{x} + \mathbf{c} + \mathbf{A}^T = \boldsymbol{\mu}$$

give

$$\mu_1 = -8,\ \mu_2 = -10.$$

The initial tableau is

x_1	x_2	x_3	s_1	t_1	u_1	u_2	u_3	v_1	v_2	v_3		θ	
3	2	1*	0	0	0	0	0	0	0	0	6		
2	0	9	3	−3	−1	0	0	1*	0	0	8	$\frac{8}{3}$	
0	2	0	2	−2	0	−1	0	0	1*	0	10	$\frac{10}{2}$	
0	0	0	1	−1	0	0	−1	0	0	1*	0	$\frac{0}{1}$	←
0	0	0	0	0	0	0	0	1	1	1	0		
−2	−2	0	−6	6	1	1	0	0	0	0	−18		
			↑										

Here the basic variables are x_3, v_1, v_2, v_3 indicated by *, and the usual simplex procedure leads to the introduction of s_1 into the basis and the removal of v_3 as indicated.

3	2	1*	0	0	0	0	0	0	0	0	6	$\frac{6}{3}$	←
2	0	0	0	0	−1	0	3	1*	0	−3	8	$\frac{8}{2}$	
0	2	0	0	0	0	−1	2	0	1*	−2	10		
0	0	0	1*	−1	0	0	−1	0	0	1	0		
−2	−2	0	0	0	1	1	−5	0	0	6	−18		
↑							× ↑						

At this stage the usual simplex procedure would choose a pivot in the u_3 column and make u_3 a basic variable with value $\frac{8}{3}$. This cannot be allowed as x_3 is a basic variable. We could make u_3 a basic variable if x_3 were the basic variable to be replaced, but that is not the case here; instead we choose a pivot in the x_1 column corresponding to the next largest negative *e.c.c.* This leads to the following tableau.

1*	$\frac{2}{3}$	$\frac{1}{3}$	0	0	0	0	0	0	0	0	2		
0	$-\frac{4}{3}$	$-\frac{2}{3}$	0	0	−1	0	3	1*	0	−3	4	$\frac{4}{3}$	←
0	2	0	0	0	0	−1	2	0	1*	−2	10	$\frac{10}{2}$	
0	0	0	1*	−1	0	0	−1	0	0	1	0		
0	$-\frac{2}{3}$	$\frac{2}{3}$	0	0	1	1	−5	0	0	6	−14		
							↑						

Again the largest negative *e.c.c.* corresponds to μ_3, but this time, as x_3 is no longer a basic variable, we can make μ_3 a basic variable.

1*	$\frac{2}{3}$	$\frac{1}{3}$	0	0	0	0	0	0	0	0	2	3	
0	$-\frac{4}{9}$	$-\frac{2}{9}$	0	0	$-\frac{1}{3}$	0	1*	$\frac{1}{3}$	0	-1	$\frac{4}{3}$		
0	$\frac{26}{9}$	$\frac{4}{9}$	0	0	$\frac{2}{3}$	-1	0	$-\frac{2}{3}$	1*	0	$\frac{22}{3}$	$\frac{33}{13}$	←
0	$-\frac{4}{9}$	$-\frac{2}{9}$	1*	1	$-\frac{1}{3}$	0	0	$\frac{1}{3}$	0	0	$\frac{4}{3}$		
0	$-\frac{26}{9}$	$-\frac{4}{9}$	0	0	$-\frac{2}{3}$	1	0	$\frac{5}{3}$	0	1	$-\frac{22}{3}$		
	↑												

1*	0	$\frac{3}{13}$	0	0	$-\frac{2}{13}$	$\frac{3}{13}$	0	$\frac{2}{3}$	$-\frac{3}{13}$	0	$\frac{4}{13}$	
0	0	$-\frac{2}{13}$	0	0	$-\frac{3}{13}$	$-\frac{2}{13}$	1*	$\frac{3}{13}$	$\frac{2}{13}$	-1	$\frac{32}{13}$	
0	1*	$\frac{2}{13}$	0	0	$\frac{3}{13}$	$-\frac{9}{26}$	0	$-\frac{3}{13}$	$\frac{9}{26}$	0	$\frac{33}{13}$	
0	0	$-\frac{2}{13}$	1*	-1	$-\frac{3}{13}$	$-\frac{2}{13}$	0	$\frac{3}{13}$	$\frac{2}{13}$	0	$\frac{32}{13}$	
0	0	0	0	0	0	0	0	1	1	1	0	

All *e.c.c.s* are now non-negative and $\mathbf{v} = \mathbf{0}$ so we have the optimum solution, which is

$$x_1 = \tfrac{4}{13},\ x_2 = \tfrac{33}{13},\ x_3 = 0,\ \mu_1 = \mu_2 = 0,$$
$$\mu_3 = \tfrac{32}{13},\ \lambda_1 = \tfrac{32}{13}.$$

It can be shown that, for $\mathbf{D}$ positive definite, the method described above cannot terminate unless the vector $\mathbf{v}$ has value zero.

The above approach to quadratic programming problems is due to P. Wolfe. For further information and other methods see {9} and {10}.

14.3 Functional Approximation

A central problem in mathematics is to approximate a given function $f(x)$ by a simpler function $p(x)$ of specified form. Typically $p(x)$ is a polynomial of degree n,

$$p(x) = p_0 + p_1 x + \ldots + p_n x^n,$$

but we may just as easily consider the more general case where, instead of a linear combination of powers of x, $p(x)$ is a linear combination of some chosen basis functions $\phi_0(x), \phi_1(x), \ldots, \phi_n(x)$.

If we regard

$$e(x) = f(x) - p(x)$$

as the error in the approximation, then $e(x)$ is a linear function of the parameters $p_0, p_1, \ldots, p_n$, and what is usually required is the best choice of $\mathbf{p} = (p_0, p_1, \ldots, p_n)^T$, namely that which makes $e(x)$ as small as possible.

The solution of such a problem depends on the way in which we measure the size of the function $e(x)$. The measures commonly used in practice are the L_1, L_2 and L_∞ function norms defined by

$$\|e(x)\|_1 = \int_a^b |e(x)|dx,$$
$$\|e(x)\|_2 = (\int_a^b (e(x))^2 dx)^{1/2}$$
$$\text{and} \quad \|e(x)\|_\infty = \max_{a \le x \le b} |e(x)|$$

respectively, where $a \le x \le b$ is the interval on which the approximation is required.

When $f(x)$ is known at only a finite number of points $x_1, x_2, \ldots, x_m$, we have a discrete approximation problem and the corresponding measures of the error are the L_1, L_2 and L_∞ vector norms $\|\mathbf{e}\|_1$, $\|\mathbf{e}\|_2$, $\|\mathbf{e}\|_\infty$ defined by

$$\Sigma_{j=1}^m |e_j|, \ (\Sigma_{j=1}^m e_j^2)^{1/2}, \ \max_{1 \le j \le m} |e_j| \quad \text{respectively,}$$

where $e_j = e(x_j), j = 1, 2, \ldots, m$. A discrete function approximation problem is often used to provide an approximate solution of the corresponding continuous function approximation problem. For further information on functional approximation see {6}, {7}, {10}.

For L_2 approximation an explicit expression for the best approximation is available in both the discrete and continuous cases. When the parameters $p_0, p_1, \ldots, p_n$ are constrained to lie in some given intervals the best approximation can be found by solving a *q.p.p.* as described in exercise 14.2.

For discrete L_1 approximation and discrete L_∞ approximation, the best approximation can be found by solving a *l.p.p.*

14.4 L_∞ Approximation

For any approximating function $p(x)$ we define

$$\max_{i=1,2,\ldots,m} |e(x_i)| = \max_{i=1,2,\ldots,m} |f(x_i) - p(x_i)| = e,$$

so that $\quad |f(x_i) - p(x_i)| \le e, \qquad i = 1, 2, \ldots, m,$

$$\text{or} \quad \left.\begin{array}{r} f(x_i) - p(x_i) \le e \\ f(x_i) - p(x_i) \ge -e \end{array}\right\}, \quad i = 1, 2, \ldots, m. \tag{1}$$

Since $p(x_i) = p_0\phi_0(x_i) + p_1\phi_1(x_i) + \ldots + p_n\phi_n(x_i)$ and e are unknown, we write the constraints (1) as

$$\left.\begin{array}{r} p(x_i) + e \ge f(x_i) \\ p(x_i) - e \le f(x_i) \end{array}\right\}, \quad i = 1, 2, \ldots, m, \tag{2}$$

and the approximation problem becomes the *l.p.p.*

minimise $e = (0, 0, \ldots, 0, 1)\begin{pmatrix}\mathbf{p}\\ e\end{pmatrix}$ *subject to the 2m linear constraints* (2).

To solve this *l.p.p.*, if we put

$$\begin{aligned}\phi_j(x_i) &= a_{ji}, &(3)\\ \bar{x}_{n+1} &= -\left(\min_{\substack{j=1,2,\ldots,n\\ p_j<0}} p_j\right) \quad \text{and}\\ \bar{x}_j &= p_j + \bar{x}_{n+1}, \quad j = 1, 2, \ldots, n,\end{aligned}$$

then $p(x)$ becomes

$$\bar{x}_0\phi_0(x) + \bar{x}_1\phi_1(x) + \ldots + \bar{x}_n\phi_n(x) - \bar{x}_{n+1}\,\Sigma_{j=0}^{n}\,\phi_j(x),$$

and the constraints (2) become

$$\begin{aligned}a_{i0}\bar{x}_0 + a_{i1}\bar{x}_1 + \ldots + a_{in}\bar{x}_n + a_{i,n+1}\bar{x}_{n+1} + e \geq f_i\\ a_{i0}\bar{x}_0 + a_{i1}\bar{x}_1 + \ldots + a_{in}\bar{x}_n + a_{i,n+1}\bar{x}_{n+1} - e \leq f_i\end{aligned} \qquad (4)$$

for $i = 1, 2, \ldots, m$, where $f_i = f(x_i)$ and

$$a_{i,n+1} = -\Sigma_{j=0}^{n}\,\phi_j(x_i). \qquad (5)$$

The $2m$ constraints (4) now involve $(n + 2)$ non-negative variables $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_n$, e so we have a *l.p.p.* in standard form if we multiply the second constraint of (4) by -1.

As we will usually have $m >> n$ we turn our attention to the dual problem, which is

$$\begin{aligned}&\textit{maximise} \quad \Sigma_{i=1}^{m}\, u_i f_i - \Sigma_{i=1}^{m}\, v_i f_i \quad \textit{subject to}\\ &\mathbf{u}^T\mathbf{A} - \mathbf{v}^T\mathbf{A} \leq \mathbf{0}, \;\; \mathbf{u}^T\mathbf{e} + \mathbf{v}^T\mathbf{e} \leq 1, \;\; \mathbf{u}, \mathbf{v} \geq \mathbf{0},\end{aligned} \qquad (6)$$

where $\mathbf{A}$ is the $m \times (n + 2)$ matrix defined by (3) and (5) and $\mathbf{e}$ is the m-vector $(1, 1, \ldots, 1)^T$.

The *l.p.p.* (6) involves $(n + 3)$ constraints in $2m$ non-negative variables but as, by the definition of $\bar{x}_{n+1}$ and $a_{i,n+1}$, the $(n + 2)$-th constraint is equal to the sum of the first $(n + 1)$ constraints multiplied by -1 we have one redundant constraint in the dual problem.

Solving this *l.p.p.* by introducing $(n + 3)$ slack variables will involve eliminating one of the $(n + 3)$ equations as described in section 4.5, so that the optimum solution of the primal will satisfy $(n + 2)$ of the primal inequality constraints as equalities. These $(n + 2)$ constraints must correspond to distinct points x_i (ER) so we see that the error in the best approximation will attain its maximum value at least $(n + 2)$ times. This corresponds to the celebrated *Chebyshev Equioscillation Theorem* which characterizes best L_∞ approximation and states, for the continuous case, that

$p(x)$ is the best approximation to $f(x)$ on $a \leq x \leq b$ in the sense of the L_∞ norm if and only if $f(x) - p(x)$ attains its maximum

magnitude at $(n+2)$ points (at least) in $a \leq x \leq b$, and that $f(x)-p(x)$ is alternately positive and negative at each pair of adjacent points.

An extensive discussion of L_∞ approximation is given in {10}, and details of an efficient algorithm in {17}.

14.5 L_1 Approximation

Again we denote $f(x_i)-p(x_i)$ by e_i, $i = 1, 2, \ldots, m$, and as e_i is a variable whose value is to be determined but which may be positive or negative, we replace it by the difference of two positive variables

$$e_i = z_i - w_i, \quad i = 1, 2, \ldots, m.$$

The objective function $\Sigma_{i=1}^m |e_i|$ will be minimised when $\Sigma_{i=1}^m (z_i + w_i)$ is minimised (*ER*).

So the L_1 approximation problem becomes

minimise $\Sigma_{i=1}^m (z_i + w_i)$

subject to $\Sigma_{j=0}^n (p_j' - q_j')\phi_j(x_i) + z_i - w_i = f_i, \quad i = 1, 2, \ldots, m$ (1)

and $p_0', p_1', \ldots, p_n', q_0', q_1', \ldots, q_n', z_1, z_2, \ldots, z_m, w_1, w_2, \ldots, w_n \geq 0$,

where the variables $p_j, j = 0, 1, \ldots, n$, have been written as the difference of two positive variables p_j', q_j'.

With $\phi_j(x_i) = a_{ji}$ the *l.p.p.* (1) may be written

$$\textit{minimise} \quad (\mathbf{0}^T, \mathbf{0}^T, \mathbf{e}^T, \mathbf{e}^T)\begin{pmatrix}\mathbf{p}'\\ \mathbf{q}'\\ \mathbf{z}\\ \mathbf{w}\end{pmatrix}$$

$$\textit{subject to} \quad (\mathbf{A}, -\mathbf{A}, \mathbf{I}_m, -\mathbf{I}_m)\begin{pmatrix}\mathbf{p}'\\ \mathbf{q}'\\ \mathbf{z}\\ \mathbf{w}\end{pmatrix} = \mathbf{b}, \quad \begin{pmatrix}\mathbf{p}'\\ \mathbf{q}'\\ \mathbf{z}\\ \mathbf{w}\end{pmatrix} \geq \mathbf{0}, \qquad (2)$$

where $b_i = f_i$, $i = 1, 2, \ldots, m$, $\mathbf{p}'$ and $\mathbf{q}'$ are $(n+1)$-vectors and $\mathbf{z}$ and $\mathbf{w}$ are m-vectors.

The dual of (2) is a bounded variable *l.p.p.* (see exercise 14.3). Although there is a modification of the simplex method to solve bounded variable problems directly (see {9}), the special form of the constraints of (2) have led to the development of an algorithm for solving the primal directly, using only the $m \times n$ array $\mathbf{A}$ of coefficients (see {18}).

As both L_1 and L_∞ approximation problems have been expressed as *l.p.p.s.*, we observe that similar problems in which there are linear constraints on the coefficients p_j of the approximating function

$$p(x) = p_0\phi_0(x) + p_1\phi_1(x) + \ldots + p_n\phi_n(x)$$

can also be solved by simply including these extra constraints in the *l.p.p*.

14.6

An interesting application of linear programming is provided by the *Wielandt-Hoffman Theorem* on the eigenvalues of symmetric matrices. If $\mathbf{A}$ is an $n \times n$ symmetric matrix ($\mathbf{A} = \mathbf{A}^T$) with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $\mathbf{B}$ is an $n \times n$ symmetric matrix with eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$ then provided that the order of the μ_i say, is suitably chosen,

$$\Sigma_{i=1}^{n} (\lambda_i - \mu_i)^2 \leq \|\mathbf{A} - \mathbf{B}\|^2, \tag{1}$$

where $\|\mathbf{A}\|$ denotes the matrix norm defined by

$$\|\mathbf{A}\|^2 = \Sigma_{i,j=1}^{n} a_{ij}^2 = trace\,(\mathbf{A}^T\mathbf{A}).$$

The most immediate application is to the situation in which $\mathbf{B} = \mathbf{A} + \delta\mathbf{A}$, and then the result gives information about the perturbation of the spectrum of eigenvalues of $\mathbf{A}$ when $\mathbf{A}$ is perturbed.

For any symmetric matrix $\mathbf{A}$ there exists an orthogonal matrix $\mathbf{Q}$ ($\mathbf{Q}^{-1} = \mathbf{Q}^T$) such that $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{A}$. For the matrix norm defined above, $\|\mathbf{Q}^T\mathbf{A}\mathbf{Q}\| = \|\mathbf{A}\|$ for any orthogonal matrix $\mathbf{Q}$, and if $\mathbf{A}$ is symmetric so is $\mathbf{Q}^T\mathbf{A}\mathbf{Q}$.

Thus if $\mathbf{Q}_A^T\mathbf{A}\mathbf{Q}_A = \mathbf{D}_\lambda$ and $\mathbf{Q}_B^T(\mathbf{Q}_A^T\mathbf{B}\mathbf{Q}_A)\mathbf{Q}_B = \mathbf{D}_\mu$, then

$$\|\mathbf{A} - \mathbf{B}\|^2 = \|\mathbf{D}_\lambda - \mathbf{Q}_B\mathbf{D}_\mu\mathbf{Q}_B^T\|^2 \geq \min_{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}} \|\mathbf{D}_\lambda - \mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T\|^2. \tag{2}$$

Now the set of orthogonal matrices $\mathbf{Q}$ includes all permutation matrices $\mathbf{P}$ (see section 3.7), and if we prove that the minimum in (2) is attained at a permutation matrix then

$\mathbf{Q}^T\mathbf{D}_\mu\mathbf{Q}$ is just $\mathbf{D}_\mu$ with its diagonal elements re-ordered, so that

$$\min_{\mathbf{Q}^T\mathbf{Q}=I} \|\mathbf{D}_\lambda - \mathbf{Q}^T\mathbf{D}_\mu\mathbf{Q}\|^2 = \Sigma_{i=1}^{n} (\lambda_i - \mu_i)^2 \tag{3}$$

and we will have established the result (1).

The problem

$$\textit{minimise} \quad \|\mathbf{D}_\lambda - \mathbf{Q}\mathbf{D}_\mu \mathbf{Q}^T\|^2 \quad \textit{subject to} \quad \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

can, surprisingly, be rewritten as a *l.p.p.*, because

$$\begin{aligned}
\|\mathbf{D}_\lambda - \mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T\|^2 &= trace\,(\mathbf{D}_\lambda - \mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T)^T(\mathbf{D}_\lambda - \mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T),\\
&= trace\,(\mathbf{D}_\lambda^T\mathbf{D}_\lambda + \mathbf{Q}\mathbf{D}_\mu^T\mathbf{D}_\mu\mathbf{Q}^T - \mathbf{D}_\lambda^T\mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T\\
&\qquad - \mathbf{Q}\mathbf{D}_\mu^T\mathbf{Q}^T\mathbf{D}_\lambda),\\
&= trace\,(\mathbf{D}_\lambda^T\mathbf{D}_\lambda + \mathbf{D}_\mu^T\mathbf{D}_\mu) + f(\mathbf{Q}),
\end{aligned}$$

$$\text{where} \quad f(\mathbf{Q}) = trace\,(-\mathbf{D}_\lambda\mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T - \mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T\mathbf{D}_\mu). \tag{4}$$

If $\mathbf{Q}$ has rows $\mathbf{q}_1^T, \mathbf{q}_2^T, \ldots, \mathbf{q}_n^T$, then the rows of $\mathbf{D}_\lambda\mathbf{Q}$ are $\lambda_1\mathbf{q}_1^T$, $\lambda_2\mathbf{q}_2^T, \ldots, \lambda_n\mathbf{q}^T$ and the columns of $\mathbf{D}_\mu\mathbf{Q}^T$ are $\mathbf{D}_\mu\mathbf{q}_1, \mathbf{D}_\mu\mathbf{q}_2, \ldots, \mathbf{D}_\mu\mathbf{q}_n$. Hence

$$\begin{aligned} f(\mathbf{Q}) &= -\{\Sigma_{i=1}^n \lambda_i\mathbf{q}_i^T\mathbf{D}_\mu\mathbf{q}_i + \Sigma_{i=1}^n \mathbf{q}_i^T\mathbf{D}_\mu\lambda_i\mathbf{q}_i\} \\ &= -\{\Sigma_{i=1}^n \Sigma_{j=1}^n \lambda_i q_{ij}\mu_j q_{ij} + \Sigma_{i=1}^n \Sigma_{j=1}^n q_{ij}\mu_j\lambda_i q_{ij}\} \\ &= -2\{\Sigma_{i,j=1}^n \lambda_i\mu_j q_{ij}^2\}. \end{aligned} \tag{5}$$

Since $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, $\Sigma_i q_{ij}^2 = \Sigma_j q_{ij}^2 = 1$, $i,j = 1,2,\ldots,n$, so writing $q_{ij}^2 = x_{ij}$ and $2\lambda_i\mu_j = r_{ij}$ we see that the problem minimise $\|\mathbf{D}_\lambda - \mathbf{Q}\mathbf{D}_\mu\mathbf{Q}^T\|^2$ over the set of matrices $\mathbf{Q}$ such that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ becomes

$$\textit{maximise } -f(\mathbf{Q}) \quad \textit{subject to } \mathbf{Q}^T\mathbf{Q} = \mathbf{I}, \text{ or}$$

$$\textit{maximise} \quad \Sigma_{i,j} r_{ij}x_{ij} \quad \textit{subject to} \quad \Sigma_i x_{ij} = \Sigma_j x_{ij} = 1, \; x_{ij} \geq 0, \quad i,j = 1,2,\ldots,n \tag{6}$$

This is precisely the marriage problem version of the transportation problem which, as we saw in chapters 10 and 12, is solved by a matrix $\mathbf{X}$ which is a permutation matrix, and so the assertion (1) is established.

For a slightly more general result on matrix eigenvalues see {19}.

Exercises 14

1. Use the *modified simplex method* described in section 14.2 to solve the quadratic programming problem

$$\begin{aligned} &\textit{minimise} \quad 3x_1^2 + 2x_2^2 + 2x_1x_2 - 18x_1 - 16x_2 \\ &\textit{subject to} \quad \text{(i) } x_1 + 2x_2 \le 10, \ x_1, x_2 \ge 0, \\ &\qquad\qquad\quad \text{(ii) } 2x_1 + x_2 \le 5, \ x_1, x_2 \ge 0. \end{aligned}$$

2. It is required to find the *best* quadratic approximation $p_1 + p_2x + p_3x^2$ to a function $g(x)$ whose values g_k at N points $x_1, x_2, \ldots, x_N$ are known, where the coefficients p_1, p_2, p_3 must satisfy

$$\begin{aligned} b_1 &\le p_1 \le b_2 \\ b_3 &\le p_2 \le b_4 \\ b_5 &\le p_3 \le b_6, \end{aligned}$$

and where *best* is to be interpreted as that which minimises the sum of squares of the residuals

$$\Sigma_{k=1}^{N} (g(x_k) - (p_1 + p_2x_k + p_3x_k^2))^2.$$

Formulate this problem as a *q.p.p.* and explain why it always has a solution.

3. Show that the *l.p.p.* (2) of section 14.5 always has an immediate initial *b.f.s.*

   Obtain the dual of this *l.p.p.* and verify that it is a *bounded variable l.p.p.* in which the usual non-negativity constraints on the variables are replaced by intervals in which they must lie.

4. For an overspecified system of linear equations

$$\mathbf{Ax} = \mathbf{b}, \quad \text{where} \quad \mathbf{A} \quad \text{is} \quad m \times n, \ m > n,$$

we cannot expect in general that there is a solution $\mathbf{x}$ satisfying the equations. Denoting the residual by $\mathbf{r}$,

$$\mathbf{Ax} - \mathbf{b} = \mathbf{r},$$

we may instead seek the best solution vector $\mathbf{x}$ in the sense of minimising $\mathbf{r}$. Formulate the *l.p.p.s* for obtaining the best solution $\mathbf{x}$

(i) using the L_1 vector norm of $\mathbf{r}$,
(ii) using the L_∞ vector norm of $\mathbf{r}$.

NOTES

APPENDIX 1

PROOF OF THEOREM 2 (Section 2.6)

Assume that the *l.p.p.* is in canonical form. Let $\mathbf{x}_0$ be a point at which $\mathbf{c}^T\mathbf{x}$ is minimised, and let $\mathbf{c}^T\mathbf{x}_0 = f_0$.

We may assume that R is bounded, because if it is not we may add to the set of constraints the constraints $x_j \leq K, j = 1, 2, \ldots, n$, where K is any sufficiently large number, e.g. $K = \max\limits_{k=1,2,\ldots,n} (\mathbf{x}_0)_k$. This will change R but not the solution of the *l.p.p.*

The point $\mathbf{x}_0$ belongs to the hyperplane H,

$$H = \{\mathbf{x} | \mathbf{c}^T\mathbf{x} = f_0\}.$$

Since H and R are closed convex sets and R is bounded,

$$T = H \cap R \quad \text{is a closed bounded convex set.}$$

Therefore we can define a sequence of sets $T, X_1, X_2, \ldots, X_n$, each closed, bounded and convex, and contained in the previous one, as follows:

$$X_1 = \{\mathbf{x}^* | x_1^* = \min_{\mathbf{x} \in T} x_1\},$$

$$X_2 = \{\mathbf{x}^* | x_2^* = \min_{\mathbf{x} \in X_1} x_2\},$$

$$\vdots$$

$$X_n = \{\mathbf{x}^* | x_n^* = \min_{\mathbf{x} \in X_{n-1}} x_n\}.$$

Now, X_n is not empty because T is not empty ($\mathbf{x}_0$ at least belongs to T). In fact, X_n contains a single point, $\mathbf{y}$ say; for suppose $\mathbf{y}$ and $\mathbf{z}$ belong to X_n, then by the definition of X_n, $y_n = z_n$. By the definition of X_{n-1}, $y_{n-1} = z_{n-1}$ and so on until $y_1 = z_1$ and thus $\mathbf{y} = \mathbf{z}$.

The point $\mathbf{y}$ is an extreme point of T (ER) and hence is an extreme point of R. For suppose $\mathbf{y}$ is not an extreme point of R, then there are $\mathbf{x}_1$ and $\mathbf{x}_2 \in R$ such that

$$\mathbf{y} = \alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2, \quad \text{where} \quad \mathbf{x}_1 \neq \mathbf{x}_2,\ 0 < \alpha < 1.$$

If $\mathbf{x}_1 \notin T$ then $\mathbf{c}^T\mathbf{x}_1 > f_0$ because $\mathbf{c}^T\mathbf{x} \geq f_0$ for any $\mathbf{x} \in R$ and $\mathbf{c}^T\mathbf{x}_1 \neq c_0$. Also $\mathbf{c}^T\mathbf{x}_1 \geq c_0$. Therefore $\mathbf{c}^T\mathbf{y} = \alpha\mathbf{c}^T\mathbf{x}_1 + (1 - \alpha)\,\mathbf{c}^T\mathbf{x}_2$

$$\text{i.e. } \mathbf{c}^T\mathbf{y} > \alpha f_0 + (1 - \alpha) f_0 = f_0,$$

which contradicts the definition of $\mathbf{x}_0$. Therefore $\mathbf{x}_1 \in T$ and similarly $\mathbf{x}_2 \in T$. But this implies that $\mathbf{y}$ is not an extreme point of T which is something we (!) have already established. Therefore $\mathbf{y}$ *is* an extreme point of R and since $\mathbf{y} \in T$, $\mathbf{c}^T\mathbf{y} = f_0$ and we have the required result.

Some comments about this rather tortuous proof are appropriate. It uses the fact that a continuous function (x_j is a continuous function of $\mathbf{x}$) attains its maximum and minimum values over a closed, bounded set at a point of the set, and so we need R to be bounded. In practice, in the simplex method, we do not need to make sure that R is bounded and so we do not need to choose K and add in the extra constraints. For this reason we do not now have to investigate whether $\mathbf{y}$ is one of the extreme points of the original R or an extreme point of the new R created by the extra constraints.

The theorem is not constructive: it does not provide us with a practical way of finding $\mathbf{x}_0$ or $\mathbf{y}$. This is partly because the definition of an extreme point of R is not a constructive one.

APPENDIX 2

DUALITY THEOREM: THIRD PROOF

We first prove the theorem of the separating hyperplane, in a somewhat abstract setting, and then use it to establish the existence of an optimum solution of the dual, given that the primal has an optimum solution. We take advantage of some simple results established during the first proof, but these do not rely on the simplex method and are easily established independently.

The Theorem of the Separating Hyperplane

Let S be a closed convex set and let $\mathbf{b}$ be any point not in S, then there is a vector $\mathbf{y}$ such that

$$\mathbf{y}^T\mathbf{b} < \inf_{\mathbf{z}\in S} \mathbf{y}^T\mathbf{z}.$$

To establish this result, let

$$\delta = \inf_{\mathbf{z}\in S} \|\mathbf{z} - \mathbf{b}\|, \text{ where the vector norm } \|\mathbf{z}\| = \|\mathbf{z}\|_2 = (\mathbf{z}^T\mathbf{z})^{1/2}.$$

As $\mathbf{b} \notin S$, $\delta > 0$. As S is closed and

$$\inf_{\mathbf{z}\in S} \|\mathbf{z} - \mathbf{b}\| = \inf_{\substack{\mathbf{z}\in S \\ \|\mathbf{z}\|\le K}} \|\mathbf{z} - \mathbf{b}\| \quad \text{for } K \text{ sufficiently large,}$$

and $\{\mathbf{z} | \mathbf{z} \in S, \|\mathbf{z}\| \le K\}$ is compact, there exists

$$\mathbf{z}_0 \in S \text{ such that } \|\mathbf{z}_0 - \mathbf{b}\| = \delta.$$

Put $\mathbf{y}_0 = \mathbf{z}_0 - \mathbf{b}$. We show that $\mathbf{y}_0$ is a satisfactory choice for $\mathbf{y}$ above. For $0 \le \alpha \le 1$ and any $\mathbf{z} \in S$,

$$\alpha\mathbf{z} + (1 - \alpha)\mathbf{z}_0 \in S, \text{ i.e. } \mathbf{z}_0 + \alpha(\mathbf{z} - \mathbf{z}_0) \in S.$$

Therefore $\|\mathbf{z}_0 + \alpha(\mathbf{z} - \mathbf{z}_0) - \mathbf{b}\|_2^2 \ge \|\mathbf{z}_0 - \mathbf{b}\|_2^2$ and therefore $2\alpha(\mathbf{z}_0 - \mathbf{b})^T(\mathbf{z} - \mathbf{z}_0) + \alpha^2(\mathbf{z} - \mathbf{z}_0)^T(\mathbf{z} - \mathbf{z}_0) \ge 0$ and considering this result as $\alpha \to 0$, we see that we must have

$$\begin{aligned} &(\mathbf{z}_0 - \mathbf{b})^T(\mathbf{z} - \mathbf{z}_0) \ge 0, \\ \text{i.e.}\quad &(\mathbf{z}_0 - \mathbf{b})^T\mathbf{z} \ge (\mathbf{z}_0 - \mathbf{b})^T\mathbf{z}_0 \\ &\qquad = (\mathbf{z}_0 - \mathbf{b})^T\mathbf{b} + (\mathbf{z}_0 - \mathbf{b})^T(\mathbf{z}_0 - \mathbf{b}) \\ &\qquad = (\mathbf{z}_0 - \mathbf{b})^T\mathbf{b} + \delta^2. \end{aligned}$$

So putting $\mathbf{z}_0 - \mathbf{b} = \mathbf{y}$, we have $\mathbf{y}^T\mathbf{z} \geq \mathbf{y}^T\mathbf{b} + \delta^2$ for any $\mathbf{z} \in S$,

$$\text{i.e.} \quad \mathbf{y}^T\mathbf{b} < \inf_{\mathbf{z}\in S} \mathbf{y}^T\mathbf{z}.$$

Exercise: prove that this result implies that

for any matrix $\mathbf{A}$ and any vector $\mathbf{b}$,
either (i) there is a vector $\mathbf{x} \geq 0$ such that $\mathbf{Ax} = \mathbf{b}$,
or (ii) there is a vector $\mathbf{y}$ such that $\mathbf{y}^T\mathbf{b} < 0$, $\mathbf{y}^T\mathbf{A} \geq \mathbf{0}^T$,
by putting $S = \{\mathbf{z}|\mathbf{Ax} = \mathbf{z},\ \mathbf{x} \geq 0\}$ and deducing that
(i) false implies (ii) true, using $\inf \mathbf{y}^T\mathbf{z} = \inf \mathbf{y}^T\mathbf{Ax} \leq \mathbf{0}$.

To establish the duality theorem, consider the *l.p.p.* in canonical primal form

$$\textit{minimise} \quad \mathbf{c}^T\mathbf{x} \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}. \tag{1}$$

Suppose that an optimum solution $\mathbf{x}_0$ exists, and put $\mathbf{c}^T\mathbf{x}_0 = f_0$. Define a set S of $(m+1)$-vectors $\mathbf{z}$ as follows:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \mathbf{z}_2 \end{pmatrix}, \quad \mathbf{z}_2 = \begin{matrix} z_2 \\ z_3 \\ \vdots \\ z_{m+1} \end{matrix},$$

$$S = \{\mathbf{z}|z_1 = tf_0 = \mathbf{c}^T\mathbf{x},\ \mathbf{z}_2 = t\mathbf{b} - \mathbf{Ax},\ t \geq 0\},$$

i.e. any $t \geq 0$ and any $\mathbf{x} \geq \mathbf{0}$ define a vector $\mathbf{z}$ in S.

Exercise: prove that S is a closed, convex cone.

We now show that $\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \notin S$, i.e. $z_1 = 1$, $\mathbf{z}_2 = \mathbf{0}$ implies that $\mathbf{z} \notin S$. Suppose $\mathbf{z}_2 = \mathbf{0} = t^*\mathbf{b} + \mathbf{Ax}^*$ and $z_1 = 1 = t^*f_0 - \mathbf{c}^T\mathbf{x}^*$, for some $t^* \geq 0$ and $\mathbf{x}^* \geq \mathbf{0}$. Then if $t^* \geq 0$, $\dfrac{1}{t^*}\mathbf{x}^*$ is feasible for the canonical primal *l.p.p.* (1) and

$$f_0 - \mathbf{c}^T\left(\frac{1}{t^*}\mathbf{x}^*\right) = \frac{1}{t^*} > 0,$$

$$\text{i.e.} \quad \mathbf{c}^T\left(\frac{1}{t^*}\mathbf{x}^*\right) < \mathbf{c}^T\mathbf{x}_0, \text{ which is a contradiction.}$$

Alternatively, if $t^* = 0$ then $\mathbf{c}^T\mathbf{x}^* = -1$ and $\mathbf{Ax}^* = \mathbf{0}$ with $\mathbf{x}^* \geq \mathbf{0}$, so $\mathbf{A}(\mathbf{x}_0 + \alpha\mathbf{x}^*) = \mathbf{b}$ and $f(\mathbf{x}_0 + \alpha\mathbf{x}^*) = f_0 - \alpha < f_0$ for $\alpha < 0$, which again contradicts the definition of $\mathbf{x}_0$.

So $\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \notin S$, and $\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}$ can take the role of $\mathbf{b}$ in the theorem of

the separating hyperplane. Therefore there exists an $(m + 1)$-vector $\begin{pmatrix}\alpha\\ \mathbf{y}\end{pmatrix}$ say, such that

$$(\alpha,\mathbf{y}^T)\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix} < (\alpha,\mathbf{y}^T)\begin{pmatrix}z_1\\ \mathbf{z}_2\end{pmatrix} \quad \text{for all} \quad \begin{pmatrix}z_1\\ \mathbf{z}_2\end{pmatrix} \in S,$$

i.e. $\alpha < \inf\limits_{\mathbf{z}\in S}(\alpha z_1 + \mathbf{y}^T\mathbf{z}_2) = \sigma$ say, and $\sigma \leq 0$ because $\begin{pmatrix}0\\ \mathbf{0}\end{pmatrix} \in S$; but if for some $\mathbf{z}^* = \begin{pmatrix}z_1^*\\ \mathbf{z}_2^*\end{pmatrix}$ which belongs to S it is true that $(\alpha,\mathbf{y}^T)\mathbf{z}^* < 0$ then $(\alpha,\mathbf{y}^T)k\mathbf{z}^*$ can be made arbitrarily large and negative by taking k sufficiently large and $k\mathbf{z}^* \in S$.

Hence $(\alpha,\mathbf{y}^T)k\mathbf{z}^* < \alpha$ for k sufficiently large, which contradicts the assertion about $(\alpha,\mathbf{y}^T)$ in the theorem of the separating hyperplane.

Therefore $\sigma \geq 0$, and therefore $\sigma = 0$. Hence any $\alpha < 0$ will suffice and we can choose $\alpha = -1$.

Thus there exists a vector $\mathbf{y}$ such that

$$(-1,\mathbf{y}^T)\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix} < (-1,\mathbf{y}^T)\mathbf{z} \quad \text{for any} \quad \mathbf{z} \in S,$$

and therefore, as S is a cone, $(-1,\mathbf{y}^T)\mathbf{z} \geq 0$ for this $\mathbf{y}$.

That is, there exists an m-vector $\mathbf{y}$ such that

$$-z_1 + \mathbf{y}^T\mathbf{z}_2 \geq 0 \quad \text{for all } \mathbf{z} \in S,$$

therefore $-tf_0 + \mathbf{c}^T\mathbf{x} + t\mathbf{y}^T\mathbf{b} - \mathbf{y}^T\mathbf{A}\mathbf{x} \geq 0$ for all $t \geq 0$ and $\mathbf{x} \geq \mathbf{0}$.

So $t(\mathbf{y}^T\mathbf{b} - f_0) + (\mathbf{c}^T - \mathbf{y}^T\mathbf{A})\mathbf{x} \geq 0$ for all $t \geq 0$ and $\mathbf{x} \geq \mathbf{0}$,

$$\text{i.e.} \quad \mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T,$$

so $\mathbf{y}$ satisfies the dual constraints, and $\mathbf{y}^T\mathbf{b} \geq f_0$.

But we know that for any $\mathbf{x}$ and $\mathbf{y}$ satisfying primal and dual constraints respectively that

$$\mathbf{y}^T\mathbf{b} \leq \mathbf{c}^T\mathbf{x}, \text{ that is } \quad \mathbf{y}^T\mathbf{b} \leq f_0 \quad \text{(see (5) of section 5.4).}$$

Therefore $\mathbf{y}^T\mathbf{b} = f_0$, so $\mathbf{y}$ is an optimum solution for the dual *l.p.p.* with the same optimum value as that of the primal.

This approach to the duality theorem follows that in {12}.

NOTES

APPENDIX 3

SOLVING SYSTEMS OF LINEAR EQUATIONS: GAUSSIAN ELIMINATION WITH INTERCHANGES; TRIANGULAR DECOMPOSITION

Throughout this book the examples are solved using exact arithmetic operations. This is very convenient for pedagogic purposes since the theoretical development assumes that this is the case. In practice however the arithmetic operations that computers perform are slightly inaccurate, for example 1 divided by 7 has an infinite decimal and binary representation and so the result cannot be stored exactly; also the product of two t-digit numbers usually has $2t$ digits and so this product cannot be stored exactly in a t-digit computer. All numbers stored in computers and the results of arithmetic operations are represented by numbers with a fixed number of digits so that input data and the results of calculations have to be *rounded off*. These arithmetic errors are equivalent to a perturbation of the problem being solved, so that given an $n \times n$ matrix $\mathbf{A}$ and an n-vector $\mathbf{b}$, whichever method is chosen to solve the system of equations $\mathbf{Ax} = \mathbf{b}$, we obtain not $\mathbf{x}$ but a *computed solution* which we call $\mathbf{x}_c$ and for which

$$\mathbf{Ax}_c \neq \mathbf{b}, \text{ but}$$
$$(\mathbf{A} + \delta\mathbf{A})\mathbf{x}_c = (\mathbf{b} + \delta\mathbf{b}),$$

where $\mathbf{x}_c$, $\delta\mathbf{A}$, $\delta\mathbf{b}$ depend on the method used as well as on $\mathbf{A}$ and $\mathbf{b}$.

An acceptable method is one which is both *efficient* in terms of the total number of arithmetic operations it requires, and *accurate* in the sense that $\delta\mathbf{A}$ and $\delta\mathbf{b}$ are small. In practice, $\delta\mathbf{A}$ and $\delta\mathbf{b}$ are unobtainable but, for any particular method, bounds for the possible magnitude of $\delta\mathbf{A}$ and $\delta\mathbf{b}$ can be found, so that a better method is one for which these bounds are smaller. Note also the stress on the difference between $(\mathbf{A},\mathbf{b})$ and the system actually solved $(\mathbf{A} + \delta\mathbf{A}, \mathbf{b} + \delta\mathbf{b})$, rather than between $\mathbf{x}$ and $\mathbf{x}_c$.

The natural way to solve $\mathbf{Ax} = \mathbf{b}$ is to systematically eliminate variables from equations by elementary row operations until the

resulting system is triangular, and then to obtain the elements of $\mathbf{x}_c$ by successive substitution.

Denoting the given system by $(\mathbf{A}^{(1)},\mathbf{b}^{(1)})$ and assuming $\mathbf{a}_{11}^{(1)} \neq 0$ we add

$(m_{i1} \times$ 1st row of $(\mathbf{A}^{(1)},\mathbf{b}^{(1)}))$ to the i-th row for $i = 2, 3, \ldots, n$,

where $\quad m_{i1} = -a_{i1}^{(1)}/a_{11}^{(1)}$.

$$\begin{aligned}
\text{Thus} \quad & a_{ij}^{(2)} = a_{ij}^{(1)} + m_{i1}a_{ij}^{(1)}, && i,j = 2,3,\ldots,n, \\
& a_{1j}^{(2)} = a_{1j}^{(1)}, && j = 1,2,\ldots,n, \\
\text{and} \quad & a_{i1}^{(2)} = 0, && i = 2,3,\ldots,n.
\end{aligned} \tag{1}$$

In general at the k-th stage we have $(\mathbf{A}^{(k)},\mathbf{b}^{(k)})$,

$$\begin{aligned}
\text{where} \quad & a_{ij}^{(k)} = 0, && j = 1,2,\ldots,k-1, i>j, \\
\text{and then} \quad & a_{ij}^{(k+1)} = a_{ij}^{(k)} + m_{ik}a_{kj}^{(k)}, && i,j = k+1,k+2,\ldots,n, \\
\text{where} \quad & m_{ik} = -a_{ik}^{(k)}/a_{kk}^{(k)}, && i = k+1,k+2,\ldots,n, \\
& a_{kj}^{(k+1)} = a_{kj}^{(k)}, && j = k, k+1,\ldots,n, \\
\text{and} \quad & a_{ik}^{(k+1)} = 0, && i = k+1,k+2,\ldots,n.
\end{aligned} \tag{2}$$

After $(n-1)$ such elimination stages we have $(\mathbf{A}^{(n)},\mathbf{b}^{(n)})$ which we denote by $(\mathbf{U},\mathbf{b}')$, where $\mathbf{U}$ is an upper triangular matrix, $u_{ij} = 0$ if $i > j$. The solution of $\mathbf{Ux} = \mathbf{b}'$ can be obtained by *back-substitution*

$$\begin{aligned}
x_n &= b'_n/u_{nn} \\
x_i &= (b'_i - \Sigma_{j=i+1}^{n} u_{ij}x_j)/u_{ii}, \quad i = n-1, n-2, \ldots, 1.
\end{aligned} \tag{3}$$

This method of obtaining $\mathbf{x}$ is called *Gaussian elimination* and as described is *not* satisfactory in general.

The elements $a_{11}^{(1)}, a_{22}^{(2)}, \ldots, a_{nn}^{(n)}$ which, at the end of the elimination, we have renamed $u_{11}, u_{22}, \ldots, u_{nn}$ are called the pivots. They play a crucial role in the process, appearing as divisors during the elimination and in the back-substitution. It is clear from (2) and (3) that the process breaks down if $a_{kk}^{(k)} = 0$, and from (3) we see that x_i can at best be as accurate as u_{ii}. If $a_{kk}^{(k)}$ is very small then its relative error due to the inexact arithmetic operations is likely to be significant, so the process must be modified to avoid small pivots if possible.

Before this, notice that if $\mathbf{L}_k$ denotes the elementary lower-triangular matrix

$$\begin{pmatrix}
1 & & & & & & \\
& 1 & & & & & \\
& & \ddots & & & & \\
& & & 1 & & & \\
& & & m_{k+1,k} & & & \\
& & & m_{k+2,k} & & \ddots & \\
& & & \vdots & & & \\
& & & m_{nk} & & & 1
\end{pmatrix} \tag{4}$$

then
$$(\mathbf{A}^{(k+1)}, \mathbf{b}^{(k+1)}) = \mathbf{L}_k(\mathbf{A}^{(k)}, \mathbf{b}^{(k)}), \tag{5}$$
and
$$\mathbf{L}_{n-1} \ldots \mathbf{L}_2\mathbf{L}_1(\mathbf{A}^{(1)}, \mathbf{b}^{(1)}) = (\mathbf{U}, \mathbf{b}').$$

Also

$$\mathbf{L}_k^{-1} = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & -m_{k+1,k} & \ddots & & \\ & & & -m_{k+2,k} & & \ddots & \\ & & & \vdots & & & \\ & & & -m_{nk} & & & 1 \end{pmatrix} \tag{6}$$

and $(\mathbf{L}_{n-1} \ldots \mathbf{L}_2\mathbf{L}_1)^{-1} = (\mathbf{L}_1^{-1}\mathbf{L}_2^{-1} \ldots \mathbf{L}_{n-1}^{-1}) = \mathbf{L}$ say, where $\mathbf{L}$ is a lower-triangular matrix whose k-th column is the k-th column of $\mathbf{L}_k^{-1}$, $k = 1, 2, \ldots, n-1$, and whose n-th column is $\mathbf{e}_n$ (*ER*). Hence $\mathbf{A}^{(1)} = \mathbf{LU}$ and $\mathbf{A}$ has been decomposed into the product of a unit lower-triangular matrix $\mathbf{L}$ and an upper-triangular matrix $\mathbf{U}$. Such a decomposition is unique, but can be obtained in other ways. Once we have such a decomposition, a system of equations can be solved directly by a forward-substitution and a back-substitution.

With $\mathbf{A} = \mathbf{LU}$,
$$\mathbf{Ax} = \mathbf{b} = \mathbf{LUx} = \mathbf{Ly} \quad \text{say.} \tag{7}$$
We obtain $\mathbf{y}$ from $\mathbf{Ly} = \mathbf{b}$, and then $\mathbf{x}$ from $\mathbf{Ux} = \mathbf{y}$.

To make the elimination process satisfactory in practice, we perform a row interchange at each stage to bring into the pivotal position the largest of the numbers $a_{ik}^{(k)}$, $i = k, k+1, \ldots, n$. Thus if
$$\max_{i=k,k+1,\ldots,n} a_{ik}^{(k)} = a_{sk}^{(k)}$$
at the k-th stage, we interchange the k-th and the s-th rows before performing the eliminations defined by (2) or (5).

This ensures that the magnitude of all multipliers m_{ij} is at most 1.

The interchange can be represented by pre-multiplication of $(\mathbf{A}^{(k)}, \mathbf{b}^{(k)})$ by a permutation matrix $\mathbf{P}_k$, so that the whole process, called *Gaussian elimination with interchanges* can be represented by
$$\mathbf{L}_{n-1}\mathbf{P}_{n-1} \ldots \mathbf{L}_2\mathbf{P}_2\mathbf{L}_1\mathbf{P}_1(\mathbf{A}^{(1)}, \mathbf{b}^{(1)}) = (\mathbf{U}, \mathbf{b}'). \tag{8}$$

Example

Here $n = 3$ and we use two-digit arithmetic throughout.

$$(\mathbf{A}^{(1)},\mathbf{b}^{(1)}) = \left(\begin{array}{ccc|c} .24 & -.32 & .18 & .10 \\ .94 & -.95 & .56 & .55 \\ -.46 & .36 & -.20 & -.30 \end{array}\right), \tag{9}$$

$$\mathbf{P}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{L}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -.26 & 1 & 0 \\ .49 & 0 & 1 \end{pmatrix},$$

$$(\mathbf{A}^{(2)},\mathbf{b}^{(2)}) = \left(\begin{array}{ccc|c} .94 & -.95 & .56 & .55 \\ 0 & -.07 & .03 & -.04 \\ 0 & -.11 & .07 & -.03 \end{array}\right),$$

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{L}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -.64 & 1 \end{pmatrix},$$

$$(\mathbf{A}^{(3)},\mathbf{b}^{(3)}) = \left(\begin{array}{ccc|c} .94 & -.95 & .56 & .55 \\ 0 & -.11 & .07 & -.04 \\ 0 & 0 & -.015 & -.021 \end{array}\right).$$

The back-substitution gives

$$x_3 = 1.4, \quad x_2 = (.04 - .07 \times 1.4)/.11 \to 1.2,$$
$$x_1 = (.55 + 95 \times 1.2 - .56 \times 1.4)/.94 \to .92.$$

The exact solution is $x_1 = x_2 = x_3 = 1.0$, so the computed solution may seem unsatisfactory. However, the system (9) is very sensitive to perturbation, and without the interchange given by $\mathbf{P}_1$ the first stage elimination with .24 as pivot and -3.9 and 1.9 as multipliers yields (in two-digit arithmetic)

$$(\mathbf{A}^{(2)},\mathbf{b}^{(2)}) = \left(\begin{array}{ccc|c} .24 & -.32 & .18 & .10 \\ 0 & .25 & -.14 & -.16 \\ 0 & -.25 & .14 & -.11 \end{array}\right), \tag{10}$$

in which $\mathbf{A}^{(2)}$ has rank 2, and the equations are inconsistent. The sensitivity of the system (9) is caused by the fact that the three equations which define $\mathbf{x}$ are independent, but *only just independent:*

$$-\mathbf{a}_{3*} + 2\,\mathbf{a}_{1*}' = (.94, -1.00, .56), \quad \mathbf{a}_{2*} = (.94, -.95, .56)$$

so that $\mathbf{A}^{(1)}$ is "nearly singular". If $a_{12}^{(1)}$ is replaced by .295, the resulting matrix is exactly singular. From a geometrical point of view, each equation represents a plane in 3-space so that the first and third equations restrict $\mathbf{x}$ to those points on their line of intersection. This line passes through the plane defined by the second equation, and so defines a unique point $\mathbf{x}$, but this line is very nearly parallel to the plane, so that a relatively small perturbation could result in a relatively large perturbation of the point $\mathbf{x}$.

We can also see a reason why large multipliers should be avoided. If the i-th row $(\mathbf{a}_{i*}, b_i)$ is replaced by

$$(\mathbf{a}_{i*}, b_i) + m(\mathbf{a}_{k*}, b_k),$$

where m is large, we will have replaced the i-th row by another row which defines a plane (hyperplane for general n) nearly parallel to that defined by the k-th row. The important information contained in the i-th row will now be in only the least significant digits of the new i-th row, and the most significant digits will be equivalent to those of the k-th row.

For example, consider $n = 3$, three-digit arithmetic and the three equations defined by

$$\left(\begin{array}{ccc|c} .00100 & .111 & .111 & .223 \\ .800 & .888 & -.888 & .800 \\ 0 & 0 & .111 & .111 \end{array}\right), \tag{11}$$

which are satisfied exactly by $x_1 = x_2 = x_3 = 1$.

With a row multiplier $m = -800$ we obtain

$$\left(\begin{array}{ccc|c} .00100 & .111 & .111 & .223 \\ 0 & -87.9 & 89.7 & 178 \\ 0 & 0 & .111 & .111 \end{array}\right) \tag{12}$$

and multiplying the second equation by $-.111/87.9$, so that we can compare it with the first equation, we obtain

$$0 \quad .111 \quad .113 \mid .224 . \tag{13}$$

The plane defined by (13) is still distinct from that defined by the first row of (11), but is now very nearly parallel, so the line defined by their intersection is much more sensitive to small perturbations than the line defined by the intersection of the two planes defined by the first two rows of (11).

The sensitivity of the system (9) is demonstrated by the small pivot $a_{33}^{(3)}$, because a small perturbation in $a_{13}^{(3)}$ could result in a much larger proportional change in $a_{33}^{(3)}$.

The 'cost' of the Gaussian elimination algorithm, in other words the amount of work or computer time the algorithm requires, clearly increases with n. To see the manner in which the cost increases with n we evaluate the number of arithmetic operations involved, for example multiplications. At the k-th stage, from (2), we need one division (to obtain $1/a_{kk}^{(k)}$). Then the new i-th row, for $i = k+1$, $k+2, \ldots, n$, requires one multiplication for the multiplier m_{ik} and $(n-k)$ multiplications $m_{ik}a_{kj}^{(k)}$, for $j = k+1,\ k+2, \ldots, n$. The total number of multiplications at the k-th stage is therefore $(n-k+1)(n-k)$ and the overall total is

$$\Sigma_{k=1}^{n-1}(n-k+1)(n-k) = \Sigma_{t=1}^{n-1}t^2 + t$$
$$= (n-1)n(2n-1)/6 + (n-1)n/2 = n^3/3 + n/3.$$

This is a cubic polynomial in n. The dominant term is $n^3/3$, compared with which quadratic and linear terms are unimportant, so we can say the elimination requires essentially $n^3/3$ multiplications. The operations which convert $\mathbf{b}^{(1)}$ to $\mathbf{b}^{(n)}$ require essentially $n^2/2$ multiplications, so does the back-substitution (ER). The corresponding numbers of additions/subtractions are also essentially $n^3/3$, $n^2/2$, $n^2/2$ (ER). Overall the number of arithmetic operations required to perform the Gaussian elimination algorithm is a cubic polynomial in n (essentially $n^3/3$) so we say it is a *polynomial-time* algorithm. Notice that for several right-hand-side vectors $\mathbf{b}$, only the operations on $\mathbf{b}$ and in the back-substitution have to be duplicated, not the elimination operations.

In the two-part simplex method, the arithmetic operations of the first stage effectively reduce an $m \times m$ submatrix of $\mathbf{A}$ to the unit matrix. This is the *Gauss-Jordan* elimination and is not generally recommended for solving $\mathbf{Ax} = \mathbf{b}$ in practice, even with interchanges, because it requires about 50% more arithmetic operations than reduction to triangular form. At the k-th stage of the Gauss-Jordan elimination multiples of the k-th row are added to rows $1, 2, \ldots, k-1$ as well as to rows $k+1,\ k+2, \ldots, n$. This is equivalent to pre-multiplication by the matrix $\mathbf{E}$ (see (14) on page 207).

$$\mathbf{E} = \begin{pmatrix} 1 & & & m_{1k} & & & \\ & \cdot & & \vdots & & & \\ & & \cdot & m_{k-1,k} & & & \\ & & & 1 & & & \\ & & & m_{k+1,k} & & & \\ & & & m_{k+2,k} & \cdot & & \\ & & & \vdots & & \cdot & \\ & & & m_{nk} & & & 1 \end{pmatrix} \tag{14}$$

where $m_{ik} = -a_{ik}^{(k)}/a_{kk}^{(k)}, i = 1, 2, \ldots, k-1, k+1, k+2, \ldots, n$. The overall result is a diagonal matrix $\mathbf{A}^{(n)}$.

If in addition we replace $(\mathbf{E})_{kk}$ by $1/a_{kk}^{(k)}$ we have exactly the matrix $\mathbf{E}_k^*$ of section 3.7, and $\mathbf{A}^{(n)}$ will be a unit matrix.

At each stage of the simplex method as described in chapter 3, once the pivotal column has been chosen (i.e. once a negative *e.c.c.* c_t' has been chosen) the pivotal element is prescribed and cannot be chosen to minimise the effects of arithmetic inaccuracies. This could lead to a seriously inaccurate solution as the example (9) shows. The example uses two-digit decimal arithmetic whereas most computers and calculators use binary arithmetic equivalent to somewhere between six-digit and fourteen-digit decimal arithmetic. The higher accuracy does mean that in some examples interchanges could be dispensed with and multipliers larger than 1 used, but it *cannot in general be relied upon* to avoid the problem of unsuitable pivots in the simplex method.

However, as we saw in section 7.3 we can regard each simplex stage as solving three $m \times m$ systems of linear equations involving one matrix of coefficients.

Writing the three systems (2) of section 7.3 as

$$\mathbf{A}\mathbf{x}_1 = \mathbf{b}_1,\ \mathbf{A}\mathbf{x}_2 = \mathbf{b}_2,\ \mathbf{A}^T\mathbf{x}_3 = \mathbf{b}_3, \tag{15}$$

if we obtain $\mathbf{L}$ and $\mathbf{U}$ such that $\mathbf{A} = \mathbf{L}\mathbf{U}$ but with $\mathbf{L}^{-1}$ and $\mathbf{U}$ actually present as in (8) with $\mathbf{A} = \mathbf{A}^{(1)}$, then $\mathbf{x}_1$ and $\mathbf{x}_2$ can both be obtained by a forward- and a back-substitution as in (7).

If $\mathbf{A} = \mathbf{L}\mathbf{U}$, then $\mathbf{A}^T = \mathbf{U}^T\mathbf{L}^T$, and $\mathbf{A}^T(\mathbf{L}^T)^{-1} = \mathbf{U}^T$.

As $(\mathbf{L}^T)^{-1} = (\mathbf{L}^{-1})^T$, and $\mathbf{P}_i^T = \mathbf{P}_i$,

$$(\mathbf{A}^T,\mathbf{b}_3)\mathbf{P}_1\mathbf{L}_1^T\mathbf{P}_2\mathbf{L}_2^T \ldots \mathbf{P}_{n-1}\mathbf{L}_{n-1}^T = (\mathbf{U}^T,\mathbf{b}_3'),$$

and so the arithmetic operations for obtaining **L** and **U** only have to be performed once.

This approach to the simplex method, incorporating an interchange strategy, is safe and satisfactory from a numerical point of view. It also involves very little extra work compared to the tableau approach because at each stage the matrix **A** is that of the previous stage, with one column changed, and it is possible to make use of the **L** and **U** we already have. These can be *updated* rather than completely re-computed as described in {12} and in section 1, chapter 2 of {8}.

For an extensive discussion of the material in this section, see any of {4}, {5}, {6}, {7}.

LIST OF THEOREMS

Theorem 1, on feasible solutions of *l.p.p.s,* is in section 2.5.

Theorem 2, on optimality at extreme points, is in section 2.5 and Appendix 1.

Theorem 3, on extreme points and *b.f.s.s,* is in section 2.8.

Theorem 4, the fundamental theorem of linear programming, is in section 2.8.

Theorem 5, on finite termination of the simplex method, is in section 3.5.

Theorem 6, on canonical and standard form, is in section 5.2.

Theorem 7, on the dual linear program, is in section 5.2.

Theorem 8, the duality theorem, is in section 5.4, 6.1, and Appendix 2.

Theorem 9, the equilibrium theorem, is in section 5.6

Theorem 10, on the separating hyperplane, is in section 6.3.

Theorem 11, on the validity of the ellipsoid algorithm, is in section 9.4.

Theorem 12, on the convergence of the ellipsoid algorithm, is in section 9.4.

Theorem 13, on maximum flows and minimum cuts, is in section 11.2.

Theorem 14, on the simple assignment problem, is in section 12.2.

Theorem 15, the integer optimum assignment problem duality theorem, is in section 12.6.

Theorem 16, on quadratic programming problems, is in section 14.1.

REFERENCES

The list below of suggested sources of further reading covers the background linear algebra, related topics in numerical analysis and non-linear optimisation and more material on linear programming. In particular, the classic book on linear programming by S. I. Gass {9} describes many more particular situations and has an extensive and truly excellent list of references to linear programming material.

{1} *Linear Algebra;* A. M. Tropper; Thomas Nelson and Sons Ltd.

{2} *Linear Algebra;* S. Lipschutz; Schaum/McGraw-Hill.

{3} *Linear Equations;* P. M. Cohn; Routledge and Kegan Paul.

{4} *Computer Solution of Linear Algebraic Systems;* G. E. Forsythe and C. B. Moler; Prentice-Hall Inc.

{5} *Introduction to Matrix Computations;* G. W. Stewart; Academic Press.

{6} *Numerical Methods;* G. Dahlquist and A. Bjorck; Prentice-Hall Inc.

{7} *A First Course in Numerical Analysis;* A. Ralston and P. Rabinowitz; McGraw-Hill.

{8} *The State of the Art in Numerical Analysis;* D. A. H. Jacobs; Academic Press.

{9} *Linear Programming;* S. I. Gass; McGraw-Hill.

{10} *Optimisation Problems;* L. Collatz and W. Wetterling; Springer-Verlag.

{11} *Optimisation Algorithms for Networks and Graphs;* E. Minieka; Marcel Dekker AG.

{12} *Introduction to Linear and Nonlinear Programming;* D. G. Luenberger; Addison-Wesley Publishing Co.

{13} *Nonlinear Programming;* M. Avriel; Prentice-Hall Inc.

{14} *Optimisation;* D. M. Greig; Longman.

{15} "Khachiyan's Linear Programming Algorithm"; B. Aspvall and R. E. Stone; Journal of Algorithms 1, March 1980, 1–13.

{16} "A Bibliography for the Ellipsoid Algorithm"; P. Wolfe; IBM Research Centre, POB 218, Yorktown Heights, NY 10598, USA.

{17} *An Improved Algorithm for Discrete Chebyshev Linear Approximation;* I. Barrodale and C. Phillips; Proc. Fourth Manitoba Conference on Numerical Math. 1974, pp 177–190.

{18} "An Improved Algorithm for Discrete l_1 Linear Approximation"; I. Barrodale and F. D. K. Roberts; SIAM J. Numer. Anal. 10, 839–848.

{19} "The Variation of the Spectrum of a Normal Matrix"; A. J. Hoffman and H. W. Wielandt; Duke Math. J. 20, 37–39.

INDEX

For important names **bold** entries indicate definitions